

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
20 September 2001 (20.09.2001)

PCT

(10) International Publication Number  
**WO 01/68882 A2**

(51) International Patent Classification<sup>7</sup>: **C12N 15/63**

(21) International Application Number: PCT/US01/07870

(22) International Filing Date: 13 March 2001 (13.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
09/525,160 14 March 2000 (14.03.2000) US

(71) Applicant (*for all designated States except US*): **TRAN-  
SKARYOTIC THERAPIES, INC.** [US/US]; 195 Albany  
Street, Cambridge, MA 02139 (US).

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): **IVANOV, Evguenii**  
[RU/US]; 390 East Foxboro Street, Sharon, MA 02067  
(US).

(74) Agents: **MYERS, P., Louis** et al.; Fish & Richardson P.C.,  
225 Franklin Street, Boston, MA 02110-2804 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM,  
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,  
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,  
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,  
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— *without international search report and to be republished  
upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.*

(54) Title: METHODS OF IMPROVING HOMOLOGOUS RECOMBINATION

(57) Abstract: The invention features a method of promoting an alteration at a selected site in a target DNA, e.g., in the chromosomal DNA of a cell. The method includes providing, at the site: (a) a double stranded DNA sequence which includes a selected DNA sequence; (b) an agent which enhances homologous recombination, e.g., a Rad52 protein or a functional fragment thereof; and (c) an agent which inhibits non-homologous end joining, e.g., an agent which inactivates Ku such as an anti-Ku antibody or a Ku-binding oligomer or polymer, and allowing the alteration to occur. The agent which inhibits non-homologous end joining, e.g., a Ku inactivating agent such as an anti-Ku antibody, is preferably provided locally. Component (a), (b), and (c) can be introduced together, which is preferred, or separately.



**WO 01/68882 A2**

## METHODS OF IMPROVING HOMOLOGOUS RECOMBINATION

### *Background of the Invention*

Current approaches to treating disease by administering therapeutic proteins include *in vitro* production of therapeutic proteins for conventional pharmaceutical delivery (e.g. intravenous, subcutaneous, or intramuscular injection) and, more recently, gene therapy.

Proteins of therapeutic interest can be produced by introducing exogenous DNA encoding the protein of therapeutic interest into appropriate cells. For example, a vector which includes exogenous DNA encoding a therapeutic protein can be introduced into cells and the encoded protein expressed. It has also been suggested that endogenous cellular genes and their expression may be modified by gene targeting. See for example, U.S. Patent No.: 5,272,071, U.S. Patent No.: 5,641,670, WO 91/06666, WO 91/06667 and WO 90/11354.

### *Summary of the Invention*

The invention is based, in part, on the use of homologous recombination between a double stranded DNA sequence and a selected target DNA, e.g., chromosomal DNA in a cell, promoted by providing an agent which enhances homologous recombination, e.g., Rad52, and an agent which inhibits non-homologous end joining, e.g., a Ku inactivating agent, in sufficiently close proximity to the DNA sequence at the targeted site. It is predicted that a higher rate of homologous recombination occurred in the presence of both Rad52 and a Ku inactivating agent than in their absence. In addition, it is predicted that gene targeting aimed at altering a targeted site in a DNA, e.g., a targeted site in the chromosomal DNA in a cell, using a selected DNA sequence as a template can be promoted by providing a Rad52 protein and a Ku inactivating agent, e.g., an anti-Ku antibody. By providing a Rad52 protein and a Ku inactivating agent in close proximity to the selected DNA sequence and the target site, a higher rate of alteration by gene targeting occurs than in the absence of a Rad52 protein and a Ku inactivating agent, e.g., an anti-Ku antibody.

Accordingly, in one aspect, the invention features, a method of promoting an alteration at a selected site in a target DNA, e.g., in the chromosomal DNA of a cell. The method includes providing, at the site: (a) a double stranded DNA sequence which includes a selected DNA sequence; (b) an agent which enhances homologous recombination, e.g., a Rad52 protein or a functional fragment thereof, or a DNA sequence which encodes Rad52 or

5 a functional fragment thereof; and (c) an agent which inhibits non-homologous end joining, e.g., an agent which inactivates Ku, and allowing the alteration to occur. In a preferred embodiment, components (a), (b), and (c) are provided, e.g., introduced into the cell, such that, at the site of an interaction between the selected DNA sequence and the target DNA, the concentration of the agent which enhances homologous recombination and of the agent which  
10 inhibits non-homologous end joining are sufficient that an alteration of the site, e.g., homologous recombination or gene correction between the selected DNA sequence and the target DNA, occurs at a higher rate than would occur in the absence of the supplied agent which enhances homologous recombination and the agent which inhibits non-homologous end joining. The agent which inhibits non-homologous end joining is preferably provided  
15 locally. Preferably the agent which inhibits non-homologous end joining is a Ku inactivating agent such as an anti-Ku antibody.

Components (a), (b), and (c) can be introduced together, which is preferred, or separately. In addition, two of the components can be introduced together and the third can be introduced separately. For example, the DNA sequence and the agent which enhances  
20 homologous recombination, e.g., Rad52, can be introduced together or the DNA sequence and the agent which inhibits non-homologous end joining, e.g., a Ku inactivating agent, can be introduced together. In another preferred embodiment, the agent which enhances homologous recombination and the agent which inhibits non-homologous end joining can be introduced together.

25 Two, or preferably all, of the components can be provided as a complex. In a preferred embodiment, the method includes contacting the target DNA, e.g., by introducing into the cell, a complex which includes: (a) a double stranded DNA sequence which includes the selected DNA sequence; (b) an agent which enhances homologous recombination, e.g., a Rad52 protein or functional fragment thereof; and (c) an agent which inhibits non-  
30 homologous end joining, e.g., a Ku inactivating agent such as an anti-Ku antibody or a Ku-binding oligomer or polymer.

In a preferred embodiment, one, or more, preferably all of the components are provided by local delivery, e.g., microinjection, and are not expressed from the target genome or another nucleic acid. In a particularly preferred embodiment, the agent which inhibits non-  
35 homologous end joining, e.g., a Ku inhibiting agent, is provided by local delivery, e.g., microinjection, and is not expressed from the target genome or another nucleic acid.

5 In a preferred embodiment, the agent which inhibits non-homologous end joining is:  
an agent which inactivates hMre11, e.g., an anti-hMre11 antibody or a hMre11-binding  
oligomer or polymer; an agent which inactivates hRad50, e.g., an anti-hRad50 antibody or a  
hRad50-binding oligomer or polymer; an agent which inactivates Nbs1, e.g., an anti-Nbs1  
antibody or a hNbs1-binding oligomer or polymer; an agent which inactivates human ligase 4  
10 (hLig4), e.g., an anti-hLig4 antibody or a hLig4-binding oligomer or polymer; an agent which  
inactivates hXrcc4, e.g., an anti-hXrcc4 antibody or a hXrcc4-binding oligomer or polymer;  
an agent which inactivates a human homolog of Rap1, e.g., an antibody to a human homolog  
of Rap1 or an oligomer or polymer which binds a human homolog of Rap1; an agent which  
inactivates a human homolog of Sir2304, e.g., an antibody to a human homolog of Sir2304 or  
15 an oligomer or polymer which binds a human homolog of Sir2304; an agent which  
inactivates Ku, e.g., an anti-Ku antibody or a Ku-binding oligomer or polymer. Any of the  
agents which inhibit non-homologous end joining can be administered alone or can be  
administered in combination with one or more of the other agents which inhibit non-  
homologous end joining.

20 In a preferred embodiment, the DNA sequence is a linear DNA sequence. In a  
preferred embodiment, the linear DNA sequence can have one or more single stranded  
overhang(s).

In a preferred embodiment, the selected DNA sequence is flanked by a targeting  
sequence. The targeting sequence is homologous to the target, e.g., homologous to DNA  
25 adjacent to the site where the target DNA is to be altered or to the site where the selected  
DNA sequence is to be integrated. Such flanking sequence can be present at one or more,  
preferably both ends of the selected DNA sequence. If two flanking sequences are present,  
one should be homologous with a first region of the target and the other should be  
homologous to a second region of the target.

30 In a preferred embodiment, the DNA sequence has one or more protruding single  
stranded end, e.g., one or both of the protruding ends are 3' ends or 5' ends.

In a preferred embodiment, the agent which enhances homologous recombination is: a  
Rad52 protein or a functional fragment thereof; a Rad51 protein or a functional fragment  
thereof; a Rad54 protein or a functional fragment thereof; or a combination thereof.

35 In a preferred embodiment, the agent which enhances homologous recombination is  
adhered to, e.g., coated on, the DNA sequence. In a preferred embodiment, the Rad52  
protein or functional fragment thereof is adhered to, e.g., coated on, the DNA sequence.



5 In a preferred embodiment, the Rad52 protein or fragment thereof is human Rad52 (hRad52).

In a preferred embodiment, the anti-Ku antibody is: an anti-Ku70 antibody; an anti-Ku80 antibody. In a preferred embodiment, the anti-Ku antibody is: a humanized antibody; a human antibody; an antibody fragment, e.g., a Fab, Fab', F(ab')<sup>2</sup> or F(v) fragment.

10 In a preferred embodiment, at least one anti-Ku antibody is covalently linked to: the DNA sequence; the Rad52 protein or fragment thereof. In another preferred embodiment, at least one anti-Ku antibody is non-covalently linked to: the DNA sequence; the Rad52 protein or fragment thereof.

In a preferred embodiment, an anti-Ku70 antibody and an anti-Ku80 antibody is  
15 provided, e.g., as components of a complex.

In a preferred embodiment, the cell is: a eukaryotic cell. In a preferred embodiment, the cell is of fungal, plant or animal origin, e.g., vertebrate origin. In a preferred embodiment, the cell is: a mammalian cell, e.g., a primary or secondary mammalian cell, e.g., a fibroblast, a hematopoietic stem cell, a myoblast, a keratinocyte, an epithelial cell, an endothelial cell, a  
20 glial cell, a neural cell, a cell comprising a formed element of the blood, a muscle cell and precursors of these somatic cells; a transformed or immortalized cell line. Preferably, the cell is a human cell. Examples of immortalized human cell line useful in the present method include, but are not limited to: a Bowes Melanoma cell (ATCC Accession No. CRL 9607), a Daudi cell (ATCC Accession No. CCL 213), a HeLa cell and a derivative of a HeLa cell  
25 (ATCC Accession Nos. CCL2 CCL2.1, and CCL 2.2), a HL-60 cell (ATCC Accession No. CCL 240), a HT1080 cell (ATCC Accession No. CCL 121), a Jurkat cell (ATCC Accession No. TIB 152), a KB carcinoma cell (ATCC Accession No. CCL 17), a K-562 leukemia cell (ATCC Accession No. CCL 243), a MCF-7 breast cancer cell (ATCC Accession No. BTH 22), a MOLT-4 cell (ATCC Accession No. 1582), a Namalwa cell (ATCC Accession No.  
30 CRL 1432), a Raji cell (ATCC Accession No. CCL 86), a RPMI 8226 cell (ATCC Accession No. CCL 155), a U-937 cell (ATCC Accession No. 1593), WI-28VA13 sub line 2R4 cells (ATCC Accession No. CLL 155), a CCRF-CEM cell (ATCC Accession No. CCL 119) and a 2780AD ovarian carcinoma cell (Van Der Blick et al., Cancer Res. 48:5927-5932, 1988), as well as heterohybridoma cells produced by fusion of human cells and cells of  
35 another species. In another embodiment, the immortalized cell line can be cell line other than a human cell line, e.g., a CHO cell line, a COS cell line.

5 In a preferred embodiment, the components, e.g., the components of a complex, are introduced into the cell by microinjection.

In one preferred embodiment, the selected DNA sequence differs from the target DNA by less than 10, 8, 6, 5, 4, 3, 2, or by a single nucleotide, e.g., by a substitution, or a  
10 deletion, or an insertion.

In a preferred embodiment, the target DNA includes a mutation, e.g., the target sequence differs from wild-type sequence by about 10, 8, 6, 5, 4, 3, 2 or by a single nucleotide. Preferably, the mutation is a point mutation, e.g., a mutation due to an insertion, deletion or a substitution.

15 In a preferred embodiment, the target DNA includes a mutation and the mutation is associated with, e.g., causes, contributes to, conditions or controls, a disease or a dysfunction. Preferably, the disease or dysfunction is: cystic fibrosis; sickle cell anemia; hemophilia A; hemophilia B; von Willebrand disease type 3; xeroderma pigmentosa; thalassaemias; Lesch-Nylan syndrome; protein C resistance; a lysosomal storage disease, e.g., Gaucher disease,  
20 Fabry disease; mucopolysaccharidosis (MPS) type 1 (Hurley-Scheie syndrome), MPS type II (Hunter syndrome), MPS type IIIA (Sanfilio A syndrome), MPS type IIIB (Sanfilio B syndrome), MPS type IIIC (Sanfilio C syndrome), MPS type IIID (Sanfilio D syndrome), MPS type IVA (Morquio A syndrome), MPS type IVB (Morquio B syndrome), MPS type VI (Maroteaux-Larry syndrome), MPS type VII (Sly syndrome).

25 In a preferred embodiment, the target DNA includes a mutation and the selected DNA sequence includes a normal wild-type sequence which can correct the mutation.

In a preferred embodiment, the target DNA includes a mutation and the mutation is in the cystic fibrosis transmembrane regulator (CFTR) gene. Preferably, the mutation is one which alters the amino acid at codon 508 of the CFTR protein coding region, e.g., the  
30 mutation is a 3 base pair in-frame deletion which eliminates a phenylalanine at codon 508 of the CFTR protein. This deletion of phenylalanine-508 in the CFTR protein is found in a high percentage of subjects having cystic fibrosis. Thus, in a preferred embodiment, a selected DNA sequence including sequence encoding phenylalanine-508 as found in the wild-type CFTR gene can be used to target and correct the mutated CFTR gene.

35 In a preferred embodiment, the target DNA includes a mutation and the mutation is in the human  $\beta$ -globin gene. Preferably, the mutation is one which alters the amino acid at the sixth codon of the  $\beta$ -globin gene, e.g., the mutation is an A to T substitution in the sixth

5 codon of the  $\beta$ -globin gene. This mutation leads to a change from glutamic acid to valine in the  $\beta$ -globin protein which is found in subjects having sickle cell anemia. Thus, in a preferred embodiment, a selected DNA which encodes a wild-type amino acid residue at codon 6, e.g., a selected DNA sequence including an A as found within the sixth codon of wild-type  $\beta$ -globin gene, can be used to target and correct the mutated  $\beta$ -globin gene.

10 In a preferred embodiment, the target DNA includes a mutation and the mutation is in the Factor VIII gene. For example, a mutation can be in exon 23, 24, and/or exon 25 of the Factor VIII gene. Preferably, the mutation is one which alters the amino acid at codon 2209 of the coding region of the Factor VIII protein coding region, e.g., the mutation is a G to A substitution in exon 24 of the Factor VIII gene which leads to a change from an arginine to a glutamine at amino acid 2209 of Factor VIII. Preferably, the mutation is one which alters the amino acid at codon 2229 of the coding region of the Factor VIII protein coding region, e.g., the mutation is a G to T substitution in exon 25 of the Factor VIII gene which leads to a change from a tryptophan to a cysteine at amino acid 2229 of Factor VIII. These mutations have been associated with moderate to severe hemophilia A. Thus, in a preferred  
20 embodiment, a selected DNA sequence including either DNA which encodes a wild-type amino acid at codon 2209 of the coding region of Factor VIII gene or DNA which encodes a wild-type amino acid at codon 2229 of the coding region of the Factor VIII gene, or both, can be used to target and correct the mutated Factor VIII gene.

In a preferred embodiment, the target DNA includes a mutation and the mutation is in  
25 the Factor IX gene. For example, in subjects having hemophilia B, most of the mutations are point mutations in the Factor IX gene. Thus, in a preferred embodiment, the selected DNA sequence can include one or more nucleotides having at least one nucleotide from the wild-type Factor IX gene, to target and correct one or more of the point mutations in the Factor IX gene associated with hemophilia B.

30 In a preferred embodiment, the target DNA includes a mutation and the mutation is in the von Willebrand factor gene. Preferably, the mutation is a single cytosine deletion in a stretch of 6 cytosines at positions 2679-2684 in exon 18 of the von Willebrand gene. This mutation is found in a significant percentage of subjects having von Willebrand disease type 3. Other mutations, e.g., point mutations, associated with von Willebrand disease type 3 can  
35 also be altered as described herein. Thus, in a preferred embodiment, a selected DNA sequence including sequences found in the wild-type von Willebrand gene, e.g., the six

5 cytosines at positions 2679-2684 of the von Willebrand gene, can be used to target and correct the mutated von Willebrand gene.

In a preferred embodiment, the target DNA includes a mutation and the mutation is in the Xeroderma pigmentosum group G (XP-G) gene. Preferably, the mutation is a deletion of a single adenine in a stretch of three adenines at positions 19-21 of a 245 base-pair exon  
10 found in the XP-G gene. This deletion leads to xeroderma pigmentosum. Thus, in a preferred embodiment, a selected DNA including the wild-type sequence of the XP-G gene, e.g., three adenines at positions 19-21 of the 245 base-pair exon, can be used to target and correct the mutated XP-G gene.

Preferably, an agent which inactivates a mismatch repair protein such as Msh2, Msh6,  
15 Msh3, Mlh1, Pms2, Mlh3, Pms1, is also provided. The agent can be included in a complex.

In another preferred embodiment, the alteration includes homologous recombination between the selected DNA sequence and the target DNA, e.g., a chromosome.

In preferred embodiment, the selected DNA sequence differs from the target DNA by  
20 more than one nucleotide, e.g., it differs from the target by a sufficient number of nucleotides such that the target, or the selected DNA sequence has an unpaired region, e.g., a loop-out region. In such an application, Msh2, Msh6, Msh3, Mlh1, Pms2, Mlh3, Pms1 can also be provided, e.g., as part of a complex.

In a preferred embodiment, the alteration includes integration of the selected sequence  
25 into the target DNA and the selected DNA is integrated such that it is in a preselected relationship with a preselected element on the target, e.g., if one is a regulatory element and the other is a sequence which encodes a protein, the regulatory element functions to regulate expression of the protein encoding sequence. Flanking sequences which promote the selected integration can be used. The selected DNA sequence can be integrated 5', 3', or within, a  
30 selected target sequence, e.g., a gene or coding sequence.

In a preferred embodiment, the alteration includes integration of the selected DNA sequence and the selected DNA sequence is a regulatory sequence, e.g., an exogenous regulatory sequence. In a preferred embodiment, the regulatory sequence includes one or more of: a promoter, an enhancer, an upstream activating sequence (UAS), a scaffold-  
35 attachment region or a transcription factor-binding site. In a preferred embodiment, the regulatory sequence includes: a regulatory sequence from a metallothionein-I gene, e.g., a mouse metallothionein-I gene, a regulatory sequence from an SV-40 gene, a regulatory



5 sequence from a cytomegalovirus gene, a regulatory sequence from a collagen gene, a regulatory sequence from an actin gene, a regulatory sequence from an immunoglobulin gene, a regulatory sequence from the HMG-CoA reductase gene, a regulatory sequence from  $\gamma$  actin gene, a regulatory sequence from transcription activator YY1 gene, a regulatory sequence from fibronectin gene, or a regulatory sequence from the EF-1 $\alpha$  gene.

10 In a preferred embodiment, the selected DNA sequence includes an exon. Preferably, the exogenous exon includes: a CAP site, the nucleotide sequence ATG, and/or encoding DNA in-frame with the targeted endogenous gene.

In a preferred embodiment, the selected DNA sequence includes a splice-donor site.

15 In a preferred embodiment, the selected DNA sequence includes an exogenous regulatory sequence which when integrated into the target functions to regulate an endogenous coding sequence. The selected DNA sequence can be integrated upstream of the coding region of an endogenous gene in the target or upstream of the endogenous regulatory sequence of an endogenous gene in the target. In another preferred embodiment, the selected DNA sequence can be integrated downstream of an endogenous gene or coding region or  
20 within an intron or an endogenous gene. In another preferred embodiment, the selected DNA sequence can be integrated such that the endogenous regulatory sequence of the endogenous gene is inactive, e.g., is wholly or partially deleted.

In a preferred embodiment, the selected DNA sequence is upstream of an endogenous gene and is linked to the second exon of the endogenous gene.

25 In a preferred embodiment, the endogenous gene encodes: a hormone, a cytokine, an antigen, an antibody, an enzyme, a clotting factor, a transport protein, a receptor, a regulatory protein, a structural protein or a transcription factor. In a preferred embodiment, the endogenous gene encodes any of the following proteins: erythropoietin, calcitonin, growth hormone, insulin, insulinotropin, insulin-like growth factors, parathyroid hormone,  $\alpha$ 2-interferon (IFNA2),  $\beta$ -interferon,  $\gamma$ -interferon, nerve growth factors, FSH $\beta$ , TGF- $\beta$ , tumor necrosis factor, glucagon, bone growth factor-2, bone growth factor-7, TSH- $\beta$ , interleukin 1, interleukin 2, interleukin 3, interleukin 6, interleukin 11, interleukin 12, CSF-granulocyte (GCSF), CSF-macrophage, CSF-granulocyte/macrophage, immunoglobulins, catalytic  
30 antibodies, protein kinase C, glucocerebrosidase, superoxide dismutase, tissue plasminogen activator, urokinase, antithrombin III, DNase,  $\alpha$ -galactosidase, tyrosine hydroxylase, blood clotting factor V, blood clotting factor VII, blood clotting factor VIII, blood clotting factor IX, blood clotting factor X, blood clotting factor XIII, apolipoprotein E, apolipoprotein A-I,

5 globins, low density lipoprotein receptor, IL-2 receptor, IL-2 antagonists,  $\alpha$ -1-antitrypsin, immune response modifiers,  $\beta$ -glucoceramidase,  $\alpha$ -iduronidase,  $\alpha$ -L-iduronidase, glucosamine-N-sulfatase,  $\alpha$ -N-acetylglucosaminidase, acetylcoenzymeA: $\alpha$ -glucosamine-N-acetyltransferase, N-acetylglucosamine-6-sulfatase,  $\beta$ -galactosidase,  $\beta$ -glucuronidase, N-acetylgalactosamine-6-sulfatase, and soluble CD4.

10 In a preferred embodiment, the endogenous gene encodes follicle stimulating hormone  $\beta$  (FSH $\beta$ ) and the selected DNA sequence includes a regulatory sequence, e.g., a regulatory sequence which differs in sequence from the regulatory sequence of the FSH $\beta$  gene. Preferably, the selected DNA sequence is flanked by a targeting sequence, e.g., such targeting sequence is present at one or more, preferably both ends of the selected DNA  
15 sequence. In a preferred embodiment, the targeting sequence is homologous to a region 5' of the FSH $\beta$  coding region (SEQ ID NO:1). In a preferred embodiment, the targeting sequence directs homologous recombination within the FSH $\beta$  coding sequence or upstream of the FSH $\beta$  coding sequence. In a preferred embodiment, the targeting sequence includes at least 20, 30, 50, 100 or 1000 contiguous nucleotides from SEQ ID NO:2, which corresponds to  
20 nucleotides -7454 to -1417 of human FSH $\beta$  sequence (numbering is relative to the translation start site), or SEQ ID NO:3, which corresponds to nucleotides -696 to -155 of human FSH $\beta$  sequence.

In a preferred embodiment, the endogenous gene encodes interferon  $\alpha$ 2 (IFN $\alpha$ 2) and the selected DNA sequence includes a regulatory sequence, e.g., a regulatory sequence which  
25 differs in sequence from the regulatory sequence of the IFN $\alpha$ 2 gene. Preferably, the selected DNA sequence is flanked by a targeting sequence, e.g., such targeting sequence is present at one or more, preferably both ends of the selected DNA sequence. In a preferred embodiment, the targeting sequence is homologous to a region 5' of the IFN $\alpha$ 2 coding region. In a preferred embodiment, the targeting sequence directs homologous recombination within a  
30 region upstream of the IFN $\alpha$ 2 coding sequence. In a preferred embodiment, the targeting sequence includes at least 20, 30, 50, 100 or 1000 contiguous nucleotides from SEQ ID NO:4, which corresponds to nucleotides -4074 to -511 of human IFN $\alpha$ 2 sequence (numbering is relative to the translation start site). For example, it can include: at least 20, 30, 50, or 100 nucleotides from SEQ ID NO:7, which corresponds to nucleotides -4074 to -  
35 3796 of human IFN $\alpha$ 2 sequence; at least 20, 30, or 50 nucleotides from SEQ ID NO:8, which corresponds to nucleotides -582 to -510 of human IFN $\alpha$ 2 sequence; at least 20, 30, 50, 100,

5 or 1000 nucleotides from SEQ ID NO:9, which corresponds to nucleotides -3795 to -583 of human IFN $\alpha$ 2 sequence.

In a preferred embodiment, the endogenous gene encodes granulocyte colony stimulating factor (GCSF) and the selected DNA sequence includes a regulatory sequence, e.g., a regulatory sequence which differs in sequence from the regulatory sequence of the GCSF gene. Preferably, the selected DNA sequence is flanked by a targeting sequence, e.g., such targeting sequence is present at one or more, preferably both ends of the selected DNA sequence. In a preferred embodiment, the targeting sequence is homologous to a region 5' of the GCSF coding region. In a preferred embodiment, the targeting sequence directs homologous recombination within the GCSF coding sequence or upstream of the GCSF coding sequence. In a preferred embodiment, the targeting sequence includes at least 20, 30, 50, 100 or 1000 contiguous nucleotides from SEQ ID NO:5, which corresponds to nucleotides -6,578 to 101 of human GCSF sequence (numbering is relative to the translation start site). For example, the target sequence can include 20, 30, 50, 100 or 1000 nucleotides from SEQ ID NO:6, which corresponds to nucleotides -6,578 to -364 of the human GCSF gene.

In another preferred embodiment, the DNA sequence includes a coding region, e.g., the selected DNA sequence encodes a protein. In a preferred embodiment, the coding region encodes: a hormone, a cytokine, an antigen, an antibody, an enzyme, a clotting factor, a transport protein, a receptor, a regulatory protein, a structural protein or a transcription factor. In a preferred embodiment, the coding region encodes any of the following proteins: erythropoietin, calcitonin, growth hormone, insulin, insulinotropin, insulin-like growth factors, parathyroid hormone,  $\alpha$ 2-interferon (IFNA2),  $\beta$ -interferon,  $\gamma$ -interferon, nerve growth factors, FSH $\beta$ , TGF- $\beta$ , tumor necrosis factor, glucagon, bone growth factor-2, bone growth factor-7, TSH- $\beta$ , interleukin 1, interleukin 2, interleukin 3, interleukin 6, interleukin 11, interleukin 12, CSF-granulocyte (GCSF), CSF-macrophage, CSF-granulocyte/macrophage, immunoglobulins, catalytic antibodies, protein kinase C, glucocerebrosidase, superoxide dismutase, tissue plasminogen activator, urokinase, antithrombin III, DNase,  $\alpha$ -galactosidase, tyrosine hydroxylase, blood clotting factor V, blood clotting factor VII, blood clotting factor VIII, blood clotting factor IX, blood clotting factor X, blood clotting factor XIII, apolipoprotein E, apolipoprotein A-I, globins, low density lipoprotein receptor, IL-2 receptor, IL-2 antagonists,  $\alpha$ -1-antitrypsin, immune response modifiers,  $\beta$ -glucoceramidase,

5  $\alpha$ -iduronidase,  $\alpha$ -L-iduronidase, glucosamine-N-sulfatase,  $\alpha$ -N-acetylglucosaminidase, acetylcoenzymeA: $\alpha$ -glucosamine-N-acetyltransferase, N-acetylglucosamine-6-sulfatase,  $\beta$ -galactosidase,  $\beta$ -glucuronidase, N-acetylgalactosamine-6-sulfatase, and soluble CD4.

In a preferred embodiment, the selected DNA sequence can be integrated into the target such that it is under the control of an endogenous regulatory element. The selected  
10 DNA can be integrated downstream of an endogenous regulatory sequence or upstream of a coding region of an endogenous gene and downstream of the endogenous regulatory sequence of the gene. In another preferred embodiment, the selected DNA can be integrated downstream of an endogenous regulatory sequence such that the coding region of the endogenous gene is inactivated, e.g., is wholly or partially deleted.

15 In a preferred embodiment, the method further includes introducing an agent which inhibits a mismatch-repair protein, e.g., Msh2, Msh6, Msh3, Mlh1, Pms2, Mlh3, Pms1, or other mismatch repair proteins, or combinations thereof. Preferably, the agent is an agent which inhibits expression of a mismatch-repair protein, e.g., the agent is an antisense RNA. In a preferred embodiment, the agent is an antibody against a mismatch-repair protein. In a  
20 preferred embodiment, the antibody against the mismatch-repair protein is covalently or non-covalently linked to the complex.

In another aspect, the invention features, a composition, e.g., a complex of components, for promoting an alteration at a target DNA, e.g., a chromosome, e.g., a target  
25 DNA described herein, using a selected DNA sequence, e.g., a selected DNA sequence described herein, as a template. The composition includes: (a) a double stranded DNA sequence which includes a selected DNA sequence; (b) an agent which enhances homologous recombination, e.g., a Rad52 protein or a functional fragment thereof; and (c) an agent which inhibits non-homologous end joining, e.g., an agent which inactivates Ku. The composition  
30 can be used, for example, to alter the target DNA sequence by integration.

In a preferred embodiment, the agent which inhibits non-homologous end joining is: an agent which inactivates hMre11, e.g., an anti-hMre11 antibody or a hMre11-binding oligomer or polymer; an agent which inactivates hRad50, e.g., an anti-hRad50 antibody or a hRad50-binding oligomer or polymer; an agent which inactivates Nbs1, e.g., an anti-Nbs1  
35 antibody or a hNbs1-binding oligomer or polymer; an agent which inactivates human ligase 4 (hLig4), e.g., an anti-hLig4 antibody or a hLig4-binding oligomer or polymer; an agent which inactivates hXrcc4, e.g., an anti-hXrcc4 antibody or a hXrcc4-binding oligomer or polymer;



5 an agent which inactivates a human homolog of Rap1, e.g., an antibody to a human homolog of Rap1 or an oligomer or polymer which binds a human homolog of Rap1; an agent which inactivates a human homolog of Sir2304, e.g., an antibody to a human homolog of Sir2304 or an oligomer or polymer which binds a human homolog of Sir2304; an agent which inactivates Ku, e.g., an anti-Ku antibody or a Ku-binding oligomer or polymer. Any of the  
10 agents which inhibit non-homologous end joining can be administered alone or can be administered in combination with one or more of the other agents which inhibit non-homologous end joining.

In a preferred embodiment, the DNA sequence is a linear DNA sequence. In a preferred embodiment, the linear DNA sequence can have one or more single stranded  
15 overhang(s).

In a preferred embodiment, the selected DNA sequence is flanked by a targeting sequence. The targeting sequence is homologous to the target, e.g., homologous to DNA adjacent to the site where the target DNA is to be altered or to the site where the selected DNA sequence is to be integrated. Such flanking sequence can be present at one or more,  
20 preferably both ends of the selected DNA sequence. If two flanking sequences are present, one should be homologous to a first region of the target and the other should be homologous to a second region of the target.

In a preferred embodiment, the DNA sequence has one or more protruding single stranded end, e.g., one or both of the protruding ends are 3' ends or 5' ends.

25 In a preferred embodiment, the agent which enhances homologous recombination is: a Rad52 protein or a functional fragment thereof; a Rad51 protein or a functional fragment thereof; a Rad54 protein or a functional fragment thereof; or a combination thereof.

In a preferred embodiment, the agent which enhances homologous recombination is adhered to, e.g., coated on, the DNA sequence. In a preferred embodiment, the Rad52  
30 protein or functional fragment thereof is adhered to, e.g., coated on, the selected DNA sequence.

In a preferred embodiment, the Rad52 protein or fragment thereof is human Rad52 (hRad52).

In a preferred embodiment, the anti-Ku antibody is: an anti-Ku70 antibody; an anti-Ku80 antibody. In a preferred embodiment, the anti-Ku antibody is: a humanized antibody; a  
35 human antibody; an antibody fragment, e.g., a Fab, Fab', F(ab')<sup>2</sup> or F(v) fragment.

5 In a preferred embodiment, at least one anti-Ku antibody is covalently linked to: the selected DNA sequence; the Rad52 protein or fragment thereof. In another preferred embodiment, at least one anti-Ku antibody is covalently linked to: the selected DNA sequence; the Rad52 protein or fragment thereof.

10 In a preferred embodiment, the composition includes an anti-Ku70 antibody and an anti-Ku80 antibody.

In a preferred embodiment, the selected DNA sequence differs from the target DNA by less than 10, 8, 6, 5, 4, 3, 2 or by a single nucleotide, e.g., a substitution, or a deletion, or an insertion.

15 In a preferred embodiment, the target DNA includes a mutation, e.g., the target sequence differs from wild-type sequence by about 10, 8, 6, 5, 4, 3, 2 or by a single nucleotide. Preferably, the mutation is a point mutation, e.g., a mutation due to an insertion, deletion or a substitution.

20 In a preferred embodiment, the target DNA includes a mutation and the mutation is associated with, e.g., causes, contributes to, conditions or controls, a disease or a dysfunction. Preferably, the disease or dysfunction is: cystic fibrosis; sickle cell anemia; hemophilia A; hemophilia B; von Willebrand disease type 3; xeroderma pigmentosa; thalassaemias; Lesch-Nylan syndrome; protein C resistance; a lysosomal storage disease, e.g., Gaucher disease, Fabry disease, mucopolysaccharidosis (MPS) type 1 (Hurley-Scheie syndrome), MPS type II  
25 (Hunter syndrome), MPS type IIIA (Sanfilio A syndrome), MPS type IIIB (Sanfilio B syndrome), MPS type IIIC (Sanfilio C syndrome), MPS type IIID (Sanfilio D syndrome), MPS type IVA (Morquio A syndrome), MPS type IVB (Morquio B syndrome), MPS type VI (Maroteaux-Larry syndrome), MPS type VII (Sly syndrome).

30 In a preferred embodiment, the target DNA includes a mutation and the selected DNA sequence includes a normal wild-type sequence which can correct the mutation.

35 In a preferred embodiment, the target DNA includes a mutation and the mutation is in the cystic fibrosis transmembrane regulator (CFTR) gene. Preferably, the mutation is one which alters the amino acid at codon 508 of the CFTR protein coding region, e.g., the mutation is a 3 base pair in-frame deletion which eliminates a phenylalanine at codon 508 of the CFTR protein. This deletion of phenylalanine-508 in the CFTR protein is found in a high percentage of subjects having cystic fibrosis. Thus, in a preferred embodiment, a selected

5 DNA sequence including sequence encoding phenylalanine-508 as found in the wild-type CFTR gene can be used to target and correct the mutated CFTR gene.

In a preferred embodiment, the target DNA includes a mutation and the mutation is in the human  $\beta$ -globin gene. Preferably, the mutation is one which alters the amino acid at the sixth codon of the  $\beta$ -globin gene, e.g., the mutation is an A to T substitution in the sixth  
10 codon of the  $\beta$ -globin gene. This mutation leads to a change from glutamic acid to valine in the  $\beta$ -globin protein which is found in subjects having sickle cell anemia. Thus, in a preferred embodiment, a selected DNA which encodes a wild-type amino acid residue at codon 6, e.g., a selected DNA sequence including an A as found within the sixth codon of wild-type  $\beta$ -globin gene, can be used to target and correct the mutated  $\beta$ -globin gene.

15 In a preferred embodiment, the target DNA includes a mutation and the mutation is in the Factor VIII gene. For example, a mutation can be in exon 23, 24, and/or exon 25 of the Factor VIII gene. Preferably, the mutation is one which alters the amino acid at codon 2209 of the coding region of the Factor VIII protein coding region, e.g., the mutation is a G to A substitution in exon 24 of the Factor VIII gene which leads to a change from an arginine to a  
20 glutamine at amino acid 2209 of Factor VIII. Preferably, the mutation is one which alters the amino acid at codon 2229 of the coding region of the Factor VIII protein coding region, e.g., the mutation is a G to T substitution in exon 25 of the Factor VIII gene which leads to a change from a tryptophan to a cysteine at amino acid 2229 of Factor VIII. These mutations have been associated with moderate to severe hemophilia A. Thus, in a preferred  
25 embodiment, a selected DNA sequence including either DNA which encodes a wild-type amino acid at codon 2209 of the coding region of Factor VIII gene, or DNA which encodes a wild-type amino acid at codon 2229 of the coding region of the Factor VIII gene, or both, can be used to target and correct the mutated Factor VIII gene.

In a preferred embodiment, the target DNA includes a mutation and the mutation is in  
30 the Factor IX gene. For example, in subjects having hemophilia B, most of the mutations are point mutations in the Factor IX gene. Thus, in a preferred embodiment, the selected DNA sequence can include one or more nucleotides having at least one nucleotide from the wild-type Factor IX gene, to target and correct one or more of the point mutations in the Factor IX gene associated with hemophilia B.

35 In a preferred embodiment, the target DNA includes a mutation and the mutation is in the von Willebrand factor gene. Preferably, the mutation is a single cytosine deletion in a stretch of six cytosines at positions 2679-2684 in exon 18 of the von Willebrand gene. This

5 mutation is found in a significant percentage of subjects having von Willebrand disease type 3. Other mutations, e.g., point mutations, associated with von Willebrand disease type 3 can also be altered as described herein. Thus, in a preferred embodiment, a selected DNA sequence including sequences found in the wild-type von Willebrand gene, e.g., the six cytosines at positions 2679-2684 in exon 18 of the von Willebrand gene, can be used to target  
10 and correct the mutated von Willebrand gene.

In a preferred embodiment, the target DNA includes a mutation and the mutation is in the Xeroderma pigmentosum group G (XP-G) gene. Preferably, the mutation is a deletion of a single adenine in a stretch of three adenines at positions 19-21 of a 245 base-pair exon found in the XP-G gene. This deletion leads to xeroderma pigmentosum. Thus, in a  
15 preferred embodiment, a selected DNA including the wild-type sequence of the XP-G gene, e.g., three adenines at positions 19-21 of the 245 base-pair exon of the XP-G gene, can be used to target and correct the mutated XP-G gene.

In another preferred embodiment, the selected DNA sequence differs from the target  
20 DNA by more than one nucleotide, e.g., it differs from the target by a sufficient number of nucleotides such that the target, or the selected DNA sequence has an unpaired region, e.g., a loop-out region. Preferably, an agent which inactivates a mismatch repair protein such as Msh2, Msh6, Msh3, Mlh1, Pms2, Mlh3, Pms1, or combinations thereof, is also included in the composition, e.g., the agent can be included in the complex.

25 In a preferred embodiment, the selected DNA sequence has a flanking sequence such that it can integrate in a preselected relationship with a preselected element on a target DNA. For example, if the selected DNA is a regulatory sequence and the target DNA encodes a protein, the flanking sequence is such that it will integrate the regulatory element so that it functions to regulate expression of the protein encoding sequence. Flanking sequences which  
30 promote the selected integration can be used. The selected DNA sequence can have a flanking sequence such that it can be integrated 5', 3' or within, a selected target sequence, e.g., a gene or coding region in the target.

In a preferred embodiment, the selected DNA sequence includes a regulatory sequence, e.g., an exogenous regulatory sequence. In a preferred embodiment, the regulatory  
35 sequence includes one or more of: a promoter, an enhancer, an UAS, a scaffold-attachment region or a transcription factor-binding site. In a preferred embodiment, the regulatory sequence includes: a regulatory sequence from a metallothionein-I gene, e.g., the mouse



5 metallothionein-I gene, a regulatory sequence from an SV-40 gene, a regulatory sequence from a cytomegalovirus gene, a regulatory sequence from a collagen gene, a regulatory sequence from an actin gene, a regulatory sequence from an immunoglobulin gene, a regulatory sequence from the HMG-CoA reductase gene, a regulatory sequence from  $\gamma$  actin gene, a regulatory sequence from transcription activator YY1 gene, a regulatory sequence  
10 from fibronectin gene, or a regulatory sequence from the EF-1 $\alpha$  gene.

In a preferred embodiment, the selected DNA sequence includes an exon. Preferably, the exogenous exon includes: a CAP site, the nucleotide sequence ATG, and/or encoding DNA in-frame with the targeted endogenous gene.

In a preferred embodiment, the selected DNA sequence includes a splice-donor site.

15 In a preferred embodiment, a composition which includes a selected DNA sequence having exogenous regulatory sequence can have a flanking sequence such that it is integrated into the target such that it functions to regulate expression of an endogenous sequence. The selected DNA can be integrated into the target upstream of the coding region of an endogenous gene or coding sequence in the target, or integrated into the target upstream of  
20 the endogenous regulatory sequence of an endogenous gene or coding sequence in the target. In another preferred embodiment, the selected DNA sequence can be integrated into the target such that the endogenous regulatory sequence of the endogenous gene is inactive, e.g., is wholly or partially deleted. The selected DNA sequence can be integrated into the target downstream of the endogenous gene or coding region, or integrated within an intron of an  
25 endogenous gene.

In a preferred embodiment, the selected DNA sequence includes a regulatory sequence, e.g., a regulatory sequence which differs in sequence from the regulatory sequence of the FSH $\beta$  gene. Preferably, the selected DNA sequence is flanked by a targeting sequence, e.g., such targeting sequence is present at one or more, preferably both ends of the selected  
30 DNA sequence. In a preferred embodiment, the targeting sequence is homologous to a region 5' of FSH $\beta$  coding region (SEQ ID NO:1). In a preferred embodiment, the targeting sequence directs homologous recombination within the FSH $\beta$  coding sequence, or upstream of the FSH $\beta$  coding sequence. In a preferred embodiment, the targeting sequence includes at least 20, 30, 50, 100 or 1000 contiguous nucleotides from SEQ ID NO:2, which corresponds to  
35 nucleotides -7454 to -1417 of human FSH $\beta$  sequence (numbering is relative to the translation start site), or SEQ ID NO:3, which corresponds to nucleotides -696 to -155 of human FSH $\beta$  sequence.

5 In a preferred embodiment, the selected DNA sequence includes a regulatory sequence, e.g., a regulatory sequence which differs in sequence from the regulatory sequence of the IFN $\alpha$ 2 gene. Preferably, the selected DNA sequence is flanked by a targeting sequence, e.g., such targeting sequence is present at one or more, preferably both ends of the selected DNA sequence. In a preferred embodiment, the targeting sequence is homologous to  
10 a region 5' of IFN $\alpha$ 2 coding region. In a preferred embodiment, the targeting sequence directs homologous recombination within a region upstream of the IFN $\alpha$ 2 coding sequence. In a preferred embodiment, the targeting sequence includes at least 20, 30, 50, 100 or 1000 contiguous nucleotides from SEQ ID NO:4, which corresponds to nucleotides -4074 to -511 of human IFN $\alpha$ 2 sequence (numbering is relative to the translation start site). For example, it  
15 can include: at least 20, 30, 50, or 100 nucleotides from SEQ ID NO:7, which corresponds to nucleotides -4074 to -3796 of human IFN $\alpha$ 2 sequence; at least 20, 30, or 50 nucleotides from SEQ ID NO:8, which corresponds to nucleotides -582 to -510 of human IFN $\alpha$ 2 sequence; at least 20, 30, 50, 100, or 1000 nucleotides from SEQ ID NO:9, which corresponds to nucleotides -3795 to -583 of human IFN $\alpha$ 2 sequence.

20 In a preferred embodiment, the selected DNA sequence includes a regulatory sequence, e.g., a regulatory sequence which differs in sequence from the regulatory sequence of the GCSF gene. Preferably, the selected DNA sequence is flanked by a targeting sequence, e.g., such targeting sequence is present at one or more, preferably both ends of the selected DNA sequence. In a preferred embodiment, the targeting sequence is homologous to  
25 a region 5' of GCSF coding region. In a preferred embodiment, the targeting sequence directs homologous recombination: within the GCSF coding sequence; upstream of the GCSF coding sequence. In a preferred embodiment, the targeting sequence includes at least 20, 30, 50, 100 or 1000 contiguous nucleotides from SEQ ID NO:5, which corresponds to nucleotides -6,578 to 101 of human GCSF sequence (numbering is relative to the translation  
30 start site). For example, the target sequence can include 20, 30, 50, 100 or 1000 nucleotides from SEQ ID NO:6, which corresponds to nucleotides -6,578 to -364 of the human GCSF gene (numbering is relative to the translation start site).

In another preferred embodiment, the DNA sequence includes a coding region, e.g.,  
35 the DNA sequence encodes a protein. In a preferred embodiment, the coding region encodes: a hormone, a cytokine, an antigen, an antibody, an enzyme, a clotting factor, a transport protein, a receptor, a regulatory protein, a structural protein or a transcription factor. In a

5 preferred embodiment, the coding region encodes any of the following proteins:

erythropoietin, calcitonin, growth hormone, insulin, insulinotropin, insulin-like growth factors, parathyroid hormone,  $\alpha$ 2-interferon (IFNA2),  $\beta$ -interferon,  $\gamma$ -interferon, nerve growth factors, FSH $\beta$ , TGF- $\beta$ , tumor necrosis factor, glucagon, bone growth factor-2, bone growth factor-7, TSH- $\beta$ , interleukin 1, interleukin 2, interleukin 3, interleukin 6, interleukin 11,  
 10 interleukin 12, CSF-granulocyte (GCSF), CSF-macrophage, CSF-granulocyte/macrophage, immunoglobulins, catalytic antibodies, protein kinase C, glucocerebrosidase, superoxide dismutase, tissue plasminogen activator, urokinase, antithrombin III, DNase,  $\alpha$ -galactosidase, tyrosine hydroxylase, blood clotting factor V, blood clotting factor VII, blood clotting factor VIII, blood clotting factor IX, blood clotting factor X, blood clotting factor  
 15 XIII, apolipoprotein E, apolipoprotein A-I, globins, low density lipoprotein receptor, IL-2 receptor, IL-2 antagonists,  $\alpha$ -1-antitrypsin, immune response modifiers,  $\beta$ -glucoceramidase,  $\alpha$ -iduronidase,  $\alpha$ -L-iduronidase, glucosamine-N-sulfatase,  $\alpha$ -N-acetylglucosaminidase, acetylcoenzymeA: $\alpha$ -glucosamine-N-acetyltransferase, N-acetylglucosamine-6-sulfatase,  $\beta$ -galactosidase,  $\beta$ -glucuronidase, N-acetylgalactosamine-6-sulfatase, and soluble CD4.

20 In a preferred embodiment, the selected DNA sequence can have a flanking sequence such that when it is integrated into the target it is under the control of an endogenous regulatory element. The selected DNA can be integrated downstream of an endogenous regulatory sequence or upstream of a coding region of an endogenous gene and downstream of the endogenous regulatory sequence of the gene. In another preferred embodiment, the  
 25 selected DNA can be integrated downstream of an endogenous regulatory sequence such that the coding region of the endogenous gene is inactive, e.g., is wholly or partially deleted.

In a preferred embodiment, the composition, e.g., the complex, is introduced into a cell. Preferably, the cell is a eukaryotic cell. In a preferred embodiment, the cell is of fungal,  
 30 plant or animal origin, e.g., vertebrate origin. In a preferred embodiment, the cell is: a mammalian cell, e.g., a primary or secondary mammalian cell, e.g., a fibroblast, a hematopoietic stem cell, a myoblast, a keratinocyte, an epithelial cell, an endothelial cell, a glial cell, a neural cell, a cell comprising a formed element of the blood, a muscle cell and precursors of these somatic cells; a transformed or immortalized cell line. Preferably, the cell  
 35 is a human cell. Examples of immortalized human cell line useful in the present method include, but are not limited to: a Bowes Melanoma cell (ATCC Accession No. CRL 9607), a Daudi cell (ATCC Accession No. CCL 213), a HeLa cell and a derivative of a HeLa cell

5 (ATCC Accession Nos. CCL2 CCL2.1, and CCL 2.2), a HL-60 cell (ATCC Accession No. CCL 240), a HT1080 cell (ATCC Accession No. CCL 121), a Jurkat cell (ATCC Accession No. TIB 152), a KB carcinoma cell (ATCC Accession No. CCL 17), a K-562 leukemia cell (ATCC Accession No. CCL 243), a MCF-7 breast cancer cell (ATCC Accession No. BTH 22), a MOLT-4 cell (ATCC Accession No. 1582), a Namalwa cell (ATCC Accession No. CRL 1432), a Rafi cell (ATCC Accession No. CCL 86), a RPMI 8226 cell (ATCC  
10 Accession No. CCL 155), a U-937 cell (ATCC Accession No. 1593), WI-28VA13 sub line 2R4 cells (ATCC Accession No. CLL 155), a CCRF-CEM cell (ATCC Accession No. CCL 119) and a 2780AD ovarian carcinoma cell (Van Der Blick et al., Cancer Res. 48:5927-5932, 1988), as well as heterohybridoma cells produced by fusion of human cells and cells of  
15 another species. In another embodiment, the immortalized cell line can be cell line other than a human cell line, e.g., a CHO cell line, a COS cell line.

In a preferred embodiment, the composition further includes an agent which inhibits a mismatch-repair protein, e.g., Msh2, Msh6, Msh3, Mlh1, Pms2, Mlh3, Pms1, or other mismatch repair proteins, or combinations thereof. Preferably, the agent is an agent which  
20 inhibits expression of a mismatch-repair protein, e.g., the agent is an antisense RNA. In a preferred embodiment, the agent is an antibody against a mismatch-repair protein. In a preferred embodiment, the antibody against the mismatch-repair protein is covalently or non-covalently linked to one or more components of the composition.

25 In another aspect, the invention features, a method of providing a protein. The method includes: providing a cell made by a method described herein, and allowing the cell to express the protein.

In a preferred embodiment: the method includes: providing a cell in which the following components have been introduced at a targeted site for alteration: (a) a double  
30 stranded DNA sequence which includes a selected DNA sequence; (b) an agent which enhances homologous recombination, e.g., a Rad52 protein or a functional fragment thereof; and (c) an agent which inhibits non-homologous end joining, e.g., a Ku inactivating agent; and allowing the cell to express the protein. Expression of the protein can occur, for example, by allowing expression of a protein encoded by the DNA, or by activating  
35 expression of the protein.

In a preferred embodiment, components (a), (b), and (c) are provided, e.g., introduced into the cell, such that, at the site of an interaction between the selected DNA sequence and



5 the target DNA, the concentrations of the agent which enhances homologous recombination and of the agent which inhibits non-homologous end joining are sufficient that an alteration of the site, e.g., homologous recombination or gene correction, between the selected DNA sequence and the target DNA, occurs at a higher rate than would occur in the absence of the supplied agent which enhances homologous recombination and the agent which inhibits non-homologous end joining. The agent which inhibits non-homologous end joining is preferably provided locally.

In a preferred embodiment, components (a), (b), and (c) can be introduced together or separately. In addition, two of the components can be introduced together and the third can be introduced separately. For example, the DNA sequence and the agent which enhances homologous recombination, e.g., Rad52, can be introduced together or the DNA sequence and the agent which inhibits non-homologous end joining, e.g., a Ku inactivating agent, can be introduced together. In another preferred embodiment, the agent which enhances homologous recombination and the agent which inhibits non-homologous end joining can be introduced together.

20 Two, or preferably all, of the components can be provided as a complex. In a preferred embodiment, the method includes contacting the target DNA, e.g., by introducing into the cell, a complex which includes: (a) a double stranded DNA sequence which includes the selected DNA sequence; (b) an agent which enhances homologous recombination, e.g., a Rad52 protein or functional fragment thereof; and (c) an agent which inhibits non-homologous end joining, e.g., an agent which inactivates Ku.

In a preferred embodiment, one, or more, preferably all of the components, are provided by local delivery, e.g., microinjection, and are not expressed from the target genome or other nucleic acid. In a particularly preferred embodiment, the agent which inhibits non-homologous end joining, e.g., a Ku-inactivating agent such as an anti-Ku antibody, is provided by local delivery, e.g., microinjection, and is not expressed from the target genome or other nucleic acid.

In a preferred embodiment, the agent which inhibits non-homologous end joining is: an agent which inactivates hMre11, e.g., an anti-hMre11 antibody or a hMre11-binding oligomer or polymer; an agent which inactivates hRad50, e.g., an anti-hRad50 antibody or a hRad50-binding oligomer or polymer; an agent which inactivates Nbs1, e.g., an anti-Nbs1 antibody or a hNbs1-binding oligomer or polymer; an agent which inactivates human ligase 4 (hLig4), e.g., an anti-hLig4 antibody or a hLig4-binding oligomer or polymer; an agent which

5 inactivates hXrcc4, e.g., an anti-hXrcc4 antibody or a hXrcc4-binding oligomer or polymer;  
an agent which inactivates a human homolog of Rap1, e.g., an antibody to a human homolog  
of Rap1 or an oligomer or polymer which binds a human homolog of Rap1; an agent which  
inactivates a human homolog of Sir2304, e.g., an antibody to a human homolog of Sir2304 or  
an oligomer or polymer which binds a human homolog of Sir2304; an agent which  
10 inactivates Ku, e.g., an anti-Ku antibody or a Ku-binding oligomer or polymer. Any of the  
agents which inhibit non-homologous end joining can be administered alone or can be  
administered in combination with one or more of the other agents which inhibit non-  
homologous end joining.

In a preferred embodiment, the DNA sequence is a linear DNA sequence. In a  
15 preferred embodiment, the linear DNA sequence can have one or more single stranded  
overhang(s).

In a preferred embodiment, the selected DNA sequence is flanked by a targeting  
sequence. The targeting sequence is homologous to the target, e.g., homologous to DNA  
adjacent to the site where the target DNA is to be altered or to the site where the selected  
20 DNA sequence is to be integrated. Such flanking sequence can be present at one or more,  
preferably both ends of the selected DNA sequence. If two flanking sequences are present  
one should be homologous with a first region of the target and the other should be  
homologous to a second region of the target.

In a preferred embodiment, the DNA sequence has one or more protruding single  
25 stranded end, e.g., one or both of the protruding ends are 3' ends or 5' ends.

In a preferred embodiment, the agent which enhances homologous recombination is: a  
Rad52 protein or a functional fragment thereof; a Rad51 protein or a functional fragment  
thereof; a Rad54 protein or a functional fragment thereof; or a combination thereof.

In a preferred embodiment, the agent which enhances homologous recombination is  
30 adhered to, e.g., coated on, the DNA sequence. In a preferred embodiment, the Rad52  
protein or functional fragment thereof is adhered to, e.g., coated, on the selected DNA  
sequence.

In a preferred embodiment, the Rad52 protein or fragment thereof is human Rad52  
(hRad52).

35 In a preferred embodiment, the anti-Ku antibody is: an anti-Ku70 antibody; an anti-  
Ku80 antibody. In a preferred embodiment, the anti-Ku antibody is: a humanized antibody; a  
human antibody; an antibody fragment, e.g., a Fab, Fab', F(ab')<sup>2</sup> or F(v) fragment.

5 In a preferred embodiment, at least one anti-Ku antibody is covalently linked to: the selected DNA sequence; the agent which enhances homologous recombination, e.g., the Rad52 protein or fragment thereof. In another preferred embodiment, at least one anti-Ku antibody is non-covalently linked to: the selected DNA sequence; the agent which enhances homologous recombination, e.g., the rad52 protein or fragment thereof.

10 In a preferred embodiment, the complex includes an anti-Ku70 antibody and an anti-Ku80 antibody provided, e.g., as components of a complex.

In a preferred embodiment, the cell is: a eukaryotic cell. In a preferred embodiment, the cell is of fungal, plant or animal origin, e.g., vertebrate origin. In a preferred embodiment, the cell is: a mammalian cell, e.g., a primary or secondary mammalian cell, e.g., a fibroblast, a hematopoietic stem cell, a myoblast, a keratinocyte, an epithelial cell, an endothelial cell, a glial cell, a neural cell, a cell comprising a formed element of the blood, a muscle cell and precursors of these somatic cells; a transformed or immortalized cell line. Preferably, the cell is a human cell. Examples of immortalized human cell line useful in the present method include, but are not limited to: a Bowes Melanoma cell (ATCC Accession No. CRL 9607), a Daudi cell (ATCC Accession No. CCL 213), a HeLa cell and a derivative of a HeLa cell (ATCC Accession Nos. CCL2 CCL2.1, and CCL 2.2), a HL-60 cell (ATCC Accession No. CCL 240), a HT1080 cell (ATCC Accession No. CCL 121), a Jurkat cell (ATCC Accession No. TIB 152), a KB carcinoma cell (ATCC Accession No. CCL 17), a K-562 leukemia cell (ATCC Accession No. CCL 243), a MCF-7 breast cancer cell (ATCC Accession No. BTH 22), a MOLT-4 cell (ATCC Accession No. 1582), a Namalwa cell (ATCC Accession No. CRL 1432), a Raji cell (ATCC Accession No. CCL 86), a RPMI 8226 cell (ATCC Accession No. CCL 155), a U-937 cell (ATCC Accession No. 1593), WI-28VA13 sub line 2R4 cells (ATCC Accession No. CLL 155), a CCRF-CEM cell (ATCC Accession No. CCL 119) and a 2780AD ovarian carcinoma cell (Van Der Blick et al., Cancer Res. 48:5927-5932, 1988), as well as heterohybridoma cells produced by fusion of human cells and cells of another species. In another embodiment, the immortalized cell line can be cell line other than a human cell line, e.g., a CHO cell line, a COS cell line.

In a preferred embodiment, the components, e.g., the components of a complex, are introduced into the cell by microinjection.

35 In a preferred embodiment, the method further includes introducing an agent which inhibits a mismatch-repair protein, e.g., Msh2, Msh6, Msh3, Mlh1, Pms2, Mlh3, Pms1, or other mismatch repair proteins or combinations thereof. Preferably, the agent is an agent

5 which inhibits expression of a mismatch-repair protein, e.g., the agent is an antisense RNA. In a preferred embodiment, the agent is an antibody against a mismatch-repair protein. In a preferred embodiment, the antibody against the mismatch-repair protein is covalently or non-covalently linked to the complex.

10 In a preferred embodiment, the protein is expressed *in vitro*. In other preferred embodiments, the cell is provided in a subject, e.g., a human, and the protein is expressed in the subject. In a preferred embodiment, the protein is expressed in a subject and the cell is autologous, allogeneic or xenogeneic. Selected DNA can be introduced into a cell *in vivo*, or the cell can be removed from the subject, the selected DNA introduced *ex vivo*, and the cell returned to the subject.

15 In a preferred embodiment, the selected DNA sequence differs from the target DNA by less than 10, 8, 6, 5, 4, 3, 2, or by a single nucleotide, e.g., a substitution, or a deletion, or an insertion.

20 In a preferred embodiment, the target DNA includes a mutation, e.g., the target sequence differs from wild-type sequence by about 10, 8, 6, 5, 4, 3, 2 or by a single nucleotide. Preferably, the mutation is a point mutation, e.g., a mutation due to an insertion, deletion or a substitution.

25 In a preferred embodiment, the target DNA includes a mutation and the mutation is associated with, e.g., causes, contributes to, conditions or controls, a disease or a dysfunction. Preferably, the disease or dysfunction is: cystic fibrosis; sickle cell anemia; hemophilia A; hemophilia B; von Willebrand disease type 3; xeroderma pigmentosa; thalassaemias; Lesch-Nylan syndrome; protein C resistance; a lysosomal disease, e.g., Gaucher disease, Fabry disease, mucopolysaccharidosis (MPS) type 1 (Hurley-Scheie syndrome), MPS type II (Hunter syndrome), MPS type IIIA (Sanfilio A syndrome), MPS type IIIB (Sanfilio B syndrome), MPS type IIIC (Sanfilio C syndrome), MPS type IIID (Sanfilio D syndrome), MPS type IVA (Morquio A syndrome), MPS type IVB (Morquio B syndrome), MPS type VI (Maroteaux-Larry syndrome), MPS type VII (Sly syndrome).

30 In a preferred embodiment, the target DNA includes a mutation and the selected DNA sequence includes a normal wild-type sequence which can correct the mutation.

35 In a preferred embodiment, the target DNA includes a mutation and the mutation is in the cystic fibrosis transmembrane regulator (CFTR) gene. Preferably, the mutation is one which alters the amino acid at codon 508 of the CFTR protein-coding region, e.g., the



5 mutation is a 3 base pair in-frame deletion which eliminates a phenylalanine at codon 508 of the CFTR protein. This deletion of phenylalanine-508 in the CFTR protein is found in a high percentage of subjects having cystic fibrosis. Thus, in a preferred embodiment, a selected DNA sequence including sequence encoding phenylalanine-508 as found in the wild-type CFTR gene can be used to target and correct the mutated CFTR gene.

10 In a preferred embodiment, the target DNA includes a mutation and the mutation is in the human  $\beta$ -globin gene. Preferably, the mutation is one which alters the amino acid at the sixth codon of the  $\beta$ -globin gene, e.g., the mutation is an A to T substitution in the sixth codon of the  $\beta$ -globin gene. This mutation leads to a change from glutamic acid to valine in the  $\beta$ -globin protein which is found in subjects having sickle cell anemia. Thus, in a  
15 preferred embodiment, a selected DNA which encodes a wild-type amino acid residue at codon 6, e.g., a selected DNA sequence including an A as found within the sixth codon of wild-type  $\beta$ -globin gene, can be used to target and correct the mutated  $\beta$ -globin gene.

In a preferred embodiment, the target DNA includes a mutation and the mutation is in the Factor VIII gene. For example, a mutation can be in exon 23, 24, and/or exon 25 of the  
20 Factor VIII gene. Preferably, the mutation is one which alters the amino acid at codon 2209 of the coding region of the Factor VIII protein coding region, e.g., the mutation is a G to A substitution in exon 24 of the Factor VIII gene which leads to a change from an arginine to a glutamine at amino acid 2209 of Factor VIII. Preferably, the mutation is one which alters the amino acid at codon 2229 of the coding region of the Factor VIII protein coding region,  
25 e.g., the mutation is a G to T substitution in exon 25 of the Factor VIII gene which leads to a change from a tryptophan to a cysteine at amino acid 2229 of Factor VIII. These mutations have been associated with moderate to severe hemophilia A. Thus, in a preferred embodiment, a selected DNA sequence including either DNA which encodes a wild-type amino acid at codon 2209 of the coding region of Factor VIII gene, or DNA which encodes a  
30 wild-type amino acid at codon 2229 of the coding region of the Factor VIII gene, or both, can be used to target and correct the mutated Factor VIII gene.

In a preferred embodiment, the target DNA includes a mutation and the mutation is in the Factor IX gene. For example, in subjects having hemophilia B, most of the mutations are point mutations in the Factor IX gene. Thus, in a preferred embodiment, the selected DNA  
35 sequence can include one or more nucleotides having at least one nucleotide from the wild-type Factor IX gene, to target and correct one or more of the point mutations in the Factor IX gene associated with hemophilia B.

5 In a preferred embodiment, the target DNA includes a mutation and the mutation is in the von Willebrand factor gene. Preferably, the mutation is a single cytosine deletion in a stretch of six cytosines at positions 2679-2684 in exon 18 of the von Willebrand gene. This mutation is found in a significant percentage of subjects having von Willebrand disease type 3. Other mutations, e.g., point mutations, associated with von Willebrand disease type 3 can  
10 also be altered as described herein. Thus, in a preferred embodiment, a selected DNA sequence including sequences found in the wild-type von Willebrand gene, e.g., the six cytosines at positions 2679-2684 in exon 18 of the von Willebrand gene, can be used to target and correct the mutated von Willebrand gene.

In a preferred embodiment, the target DNA includes a mutation and the mutation is in  
15 the Xeroderma pigmentosum group G (XP-G) gene. Preferably, the mutation is a deletion of a single adenine in a stretch of adenines at positions 19-21 of a 245 base-pair exon found in the XP-G gene. This deletion leads to xeroderma pigmentosum. Thus, in a preferred embodiment, a selected DNA including the wild-type sequence of XP-G gene, e.g., three adenines at positions 19-21 at the 245 base-pair exon of the XP-G gene, can be used to target  
20 and correct the mutated XP-G gene.

In another preferred embodiment, the alteration includes homologous recombination between the selected DNA sequence and the target DNA, e.g., a chromosome.

In preferred embodiment, the selected DNA sequence differs from the target DNA by  
25 more than one nucleotide, e.g., it differs from the target by a sufficient number of nucleotides such that the target, or the selected DNA sequence has an unpaired region, e.g., a loop-out region. In such an application, Msh2, Msh6, Msh3, Mlh1, Pms2, Mlh3, Pms1, or combinations thereof, can also be provided, e.g., as part of a complex.

In a preferred embodiment, the alteration includes integration of the selected sequence  
30 into the target DNA and the selected DNA is integrated such that it is in a preselected relationship with a preselected element on the target, e.g., if one is a regulatory element and the other is a sequence which encodes a protein, the regulatory element functions to control expression of the protein encoding sequence. Flanking sequences which promote the selected integration can be used. The selected DNA sequence can be integrated 5', 3', or within, a  
35 selected target sequence, e.g., a gene or coding sequence.

In a preferred embodiment, the alteration includes integration of the selected DNA sequence and the selected DNA sequence is a regulatory sequence, e.g., an exogenous

5 regulatory sequence. In a preferred embodiment, the regulatory sequence includes one or more of: a promoter, an enhancer, an UAS, a scaffold-attachment region or a transcription factor-binding site. In a preferred embodiment, the regulatory sequence includes: a regulatory sequence from metallothionein-I gene, e.g., the mouse metallothionein gene, a regulatory sequence from an SV-40 gene, a regulatory sequence from a cytomegalovirus  
10 gene, a regulatory sequence from a collagen gene, a regulatory sequence from an actin gene, a regulatory sequence from an immunoglobulin gene, a regulatory sequence from the HMG-CoA reductase gene, a regulatory sequence from  $\gamma$  actin gene, a regulatory sequence from transcription activator YY1 gene, a regulatory sequence from fibronectin gene, or a regulatory sequence from the EF-1 $\alpha$  gene.

15 In a preferred embodiment, the selected DNA sequence includes an exon. Preferably, the exogenous exon includes: a CAP site, the nucleotide sequence ATG, and/or encoding DNA in-frame with the targeted endogenous gene.

In a preferred embodiment, the selected DNA sequence includes a splice-donor site.

In a preferred embodiment, the selected DNA sequence includes an exogenous  
20 regulatory sequence which when integrated into the target functions to regulate expression of an endogenous gene. The selected DNA can be integrated upstream of the coding region of an endogenous gene in the target or upstream of the endogenous regulatory sequence of an endogenous gene or coding region in the target. In another preferred embodiment, the selected DNA can be integrated downstream of an endogenous gene or coding region or  
25 within an intron or endogenous gene. In another preferred embodiment, the endogenous regulatory sequence of the endogenous gene is inactive, e.g., is wholly or partially deleted.

In a preferred embodiment, the selected DNA sequence is upstream of the endogenous gene and is linked to the second exon of the endogenous gene.

In a preferred embodiment, the endogenous gene encodes: a hormone, a cytokine, an  
30 antigen, an antibody, an enzyme, a clotting factor, a transport protein, a receptor, a regulatory protein, a structural protein or a transcription factor. In a preferred embodiment, the endogenous gene encodes any of the following proteins: erythropoietin, calcitonin, growth hormone, insulin, insulinotropin, insulin-like growth factors, parathyroid hormone,  $\alpha$ 2-interferon (IFNA2),  $\beta$ -interferon,  $\gamma$ -interferon, nerve growth factors, FSH $\beta$ , TGF- $\beta$ , tumor  
35 necrosis factor, glucagon, bone growth factor-2, bone growth factor-7, TSH- $\beta$ , interleukin 1, interleukin 2, interleukin 3, interleukin 6, interleukin 11, interleukin 12, CSF-granulocyte (GCSF), CSF-macrophage, CSF-granulocyte/macrophage, immunoglobulins, catalytic

5 antibodies, protein kinase C, glucocerebrosidase, superoxide dismutase, tissue plasminogen activator, urokinase, antithrombin III, DNase,  $\alpha$ -galactosidase, tyrosine hydroxylase, blood clotting factor V, blood clotting factor VII, blood clotting factor VIII, blood clotting factor IX, blood clotting factor X, blood clotting factor XIII, apolipoprotein E, apolipoprotein A-I, globins, low density lipoprotein receptor, IL-2 receptor, IL-2 antagonists,  $\alpha$ -1-antitrypsin,  
 10 immune response modifiers,  $\beta$ -glucoceramidase,  $\alpha$ -iduronidase,  $\alpha$ -L-iduronidase, glucosamine-N-sulfatase,  $\alpha$ -N-acetylglucosaminidase, acetylcoenzymeA: $\alpha$ -glucosamine-N-acetyltransferase, N-acetylglucosamine-6-sulfatase,  $\beta$ -galactosidase,  $\beta$ -glucuronidase, N-acetylgalactosamine-6-sulfatase, and soluble CD4.

In a preferred embodiment, the endogenous gene encodes follicle stimulating  
 15 hormone  $\beta$  (FSH $\beta$ ) and the selected DNA sequence includes a regulatory sequence, e.g., a regulatory sequence which differs in sequence from the regulatory sequence of the FSH $\beta$  gene. Preferably, the selected DNA sequence is flanked by a targeting sequence, e.g., such targeting sequence is present at one or more, preferably both ends of the selected DNA  
 20 sequence. In a preferred embodiment, the targeting sequence is homologous to a region 5' of FSH $\beta$  coding region (SEQ ID NO:1). In a preferred embodiment, the targeting sequence directs homologous recombination within the FSH $\beta$  coding sequence or upstream of the FSH $\beta$  coding sequence. In a preferred embodiment, the targeting sequence includes at least 20, 30, 50, 100 or 1000 contiguous nucleotides from SEQ ID NO:2, which corresponds to nucleotides -7454 to -1417 of human FSH $\beta$  sequence (numbering is relative to the  
 25 translation start site), or SEQ ID NO:3, which corresponds to nucleotides -696 to -155 of human FSH $\beta$  sequence.

In a preferred embodiment, the endogenous gene encodes interferon  $\alpha$ 2 (IFN $\alpha$ 2) and the selected DNA sequence includes a regulatory sequence, e.g., a regulatory sequence which differs in sequence from the regulatory sequence of the IFN $\alpha$ 2 gene. Preferably, the selected  
 30 DNA sequence is flanked by a targeting sequence, e.g., such targeting sequence is present at one or more, preferably both ends of the selected DNA sequence. In a preferred embodiment, the targeting sequence is homologous to a region 5' of IFN $\alpha$ 2 coding region. In a preferred embodiment, the targeting sequence directs homologous recombination within a region upstream of the IFN $\alpha$ 2 coding sequence. In a preferred embodiment, the targeting sequence  
 35 includes at least 20, 30, 50, 100 or 1000 contiguous nucleotides from SEQ ID NO:4, which corresponds to nucleotides -4074 to -511 of human IFN $\alpha$ 2 sequence (numbering is relative to the translation start site). For example, it can include: at least 20, 30, 50, or 100



5 nucleotides from SEQ ID NO:7, which corresponds to nucleotides -4074 to -3796 of human IFN $\alpha$ 2 sequence; at least 20, 30, or 50 nucleotides from SEQ ID NO:8, which corresponds to nucleotides -582 to -510 of human IFN $\alpha$ 2 sequence; at least 20, 30, 50, 100, or 1000 nucleotides from SEQ ID NO:9, which corresponds to nucleotides -3795 to -583 of human IFN $\alpha$ 2 sequence.

10 In a preferred embodiment, the endogenous gene encodes granulocyte colony stimulating factor (GCSF) and the selected DNA sequence includes a regulatory sequence, e.g., a regulatory sequence which differs in sequence from the regulatory sequence of the GCSF gene. Preferably, the selected DNA sequence is flanked by a targeting sequence, e.g., such targeting sequence is present at one or more, preferably both ends of the selected DNA  
15 sequence. In a preferred embodiment, the targeting sequence is homologous to a region 5' of GCSF coding region. In a preferred embodiment, the targeting sequence directs homologous recombination within the GCSF coding sequence or upstream of the GCSF coding sequence. In a preferred embodiment, the targeting sequence includes at least 20, 30, 50, 100 or 1000 contiguous nucleotides from SEQ ID NO:5, which corresponds to nucleotides -6,578 to 101  
20 of human GCSF sequence (numbering is relative to the translation start site). For example, the target sequence can include 20, 30, 50, 100 or 1000 nucleotides from SEQ ID NO:6, which corresponds to nucleotides -6,578 to -364 of the human GCSF gene (numbering is relative to the translation start site).

25 In another preferred embodiment, the DNA sequence includes a coding region, e.g., the DNA sequence encodes a protein. In a preferred embodiment, the coding region encodes: a hormone, a cytokine, an antigen, an antibody, an enzyme, a clotting factor, a transport protein, a receptor, a regulatory protein, a structural protein or a transcription factor. In a preferred embodiment, the coding region encodes any of the following proteins:

30 erythropoietin, calcitonin, growth hormone, insulin, insulinotropin, insulin-like growth factors, parathyroid hormone,  $\beta$ -interferon,  $\gamma$ -interferon, nerve growth factors, FSH $\beta$ , TGF- $\beta$ , tumor necrosis factor, glucagon, bone growth factor-2, bone growth factor-7, TSH- $\beta$ , interleukin 1, interleukin 2, interleukin 3, interleukin 6, interleukin 11, interleukin 12, CSF-granulocyte, CSF-macrophage, CSF-granulocyte/macrophage, immunoglobulins, catalytic  
35 antibodies, protein kinase C, glucocerebrosidase, superoxide dismutase, tissue plasminogen activator, urokinase, antithrombin III, DNase,  $\alpha$ -galactosidase, tyrosine hydroxylase, blood clotting factor V, blood clotting factor VII, blood clotting factor VIII, blood clotting factor

5 IX, blood clotting factor X, blood clotting factor XIII, apolipoprotein E, apolipoprotein A-I, globins, low density lipoprotein receptor, IL-2 receptor, IL-2 antagonists,  $\alpha$ -1-antitrypsin, immune response modifiers,  $\beta$ -glucoceramidase,  $\alpha$ -iduronidase,  $\alpha$ -L-iduronidase, glucosamine-N-sulfatase,  $\alpha$ -N-acetylglucosaminidase, acetylcoenzymeA: $\alpha$ -glucosamine-N-acetyltransferase, N-acetylglucosamine-6-sulfatase,  $\beta$ -galactosidase,  $\beta$ -glucuronidase, N-  
10 acetylgalactosamine-6-sulfatase, and soluble CD4.

In a preferred embodiment, the selected DNA sequence can be integrated into the target downstream of an endogenous regulatory sequence or upstream of a coding region of an endogenous gene and downstream of the endogenous regulatory sequence of the gene. In another preferred embodiment, the selected DNA sequence can be integrated downstream of  
15 an endogenous regulatory sequence such that the coding region of the endogenous gene is inactive, e.g., is deleted.

In another aspect, the invention features, a cell made by any of the methods described herein.

20 In another aspect, the invention features a method of altering expression of a protein coding sequence of a gene in a cell, by any of the methods described herein.

In a preferred embodiment, the method includes introducing a complex described herein having a DNA sequence which includes a regulatory sequence into the cell;  
25 maintaining the cell under conditions which permit alteration of a targeted genomic sequence to produce a homologously recombinant cell; and maintaining the homologously recombinant cell under conditions which permit expression of the protein coding sequence of the gene under control of the regulatory sequence.

30 maintaining the homologously recombinant cell under conditions which permit expression of the protein coding sequence of the gene under control of the regulatory sequence, thereby altering expression of the protein coding sequence of the gene.

The term "homologous" as used herein, refers to a targeting sequence that is identical to or sufficiently similar to a target site, e.g., a chromosomal DNA target site, so that the  
35 targeting sequence and the target site can undergo homologous recombination. A small percentage of base pair mismatches is acceptable, as long as homologous recombination can occur at a useful frequency.

5 As used herein, the term “wild-type” refers to a sequence which is not associated with, e.g., causes, contributes to, conditions or controls, a disease or dysfunction.

As used herein, a “complex” refers to a stable association in which the components are coupled by covalent or non-covalent bonds.

10 Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

### 15 *Detailed Description of the Invention*

#### Agents Which Enhance Homologous Recombination

Agents which enhance homologous recombination can be provided with a selected DNA sequence in order to promote homologous recombination between the selected DNA sequence and the target DNA, e.g., chromosomal DNA. Agents which enhance homologous recombination have one or more of the following functions: 1) increase homologous recognition between the selected DNA sequence and the selected site for integration; 2) increase homologous pairing between the selected DNA sequence and the selected site for integration; 3) increase efficiency of strand invasion and strand exchange between the recombining DNA sequences; 4) increase efficiency of processing of intermediate structures into mature products of recombination.

An agent which enhances homologous recombination can be introduced to a cell in a mixture which includes the double stranded DNA sequence, it can be introduced immediately prior to or after administration of the DNA sequence or it can be adhered, e.g., coated, on the DNA sequence. The entire DNA sequence can be coated with an agent which enhances homologous recombination, e.g., Rad52, e.g., hRad52, or a fragment thereof, or one or more of the ends of the DNA sequence can be coated, e.g., one or more of a protruding single stranded end of the DNA sequence can be coated. Preferably, the agent which enhances homologous recombination coats at least a portion of a protruding single stranded 3' end or 5' end of the DNA sequence.

5           Examples of agents which enhance homologous recombination include: Rad52 or a functional fragment thereof; Rad51 or a functional fragment thereof; Rad54 or a functional fragment thereof; or a combination or two or more of these proteins or fragments of these proteins. The agent which enhances homologous recombination can also be expressed intercellularly, e.g., a nucleic acid sequence encoding any of the above-described agents can  
10 be introduced into a cell.

          A determination of whether a Rad51 fragment is functional can be made by known techniques. For example, the functionality of a Rad51 fragment can be determined based on its ability to mediate homologous pairing and strand exchange in an *in vitro* assay known in the art, e.g., as described in Baumann et al. (1996) *Cell* 87:757-766. Briefly, hRad51 is first  
15 preincubated with circular ssDNA and then <sup>32</sup>P-labeled linear duplex DNA is added. The formation of joint molecules and the amount of strand exchange can be determined by electrophoresis. In addition, the functionality of a Rad51 fragment can be determined based on its ability to bind nicked duplex DNA in the presence of ATP to form helical nucleoprotein filament which can be visualized by electron microscopy as described in  
20 Benson et al. (1994) *EMBO J.* 13:5764-5771. The functionality of Rad51 can also be determined based on its ability to alleviate defects in DNA repair and homologous recombination in cells lacking functional Rad51 protein. Thus, it can be determined if a Rad51 fragment is functional if it confers a positive effect in the above-mentioned assays as compared to its absence. Moreover, the extent of the positive effect conferred by a Rad51  
25 fragment can be compared to the extent of positive effect conferred by full-length Rad51.

          The functionality of a Rad54 fragment can be determined based on its ability to hydrolyze ATP in the presence of dsDNA in an assay known in the art, e.g., as described in Swagemakers et al. (1998) *J. Biol. Chem.* 273:28292-28297. In addition, the functionality of a Rad54 fragment can be determined based on its ability to alleviate defects in DNA repair  
30 and homologous recombination in cells lacking functional Rad54 protein.

#### Rad52 and Functional Fragments Thereof

          Rad52 provided with a DNA sequence at a selected site in a target DNA, e.g., a  
35 selected site in chromosomal DNA, can provide a higher rate of alteration of the site, e.g., homologous recombination, than would occur in its absence. While not wishing to be bound



5 by theory, it is believed that Rad52 can provide one or more of the following functions: 1) protect the entire DNA sequence from nuclease degradation; 2) protect a protruding single stranded end of the DNA sequence, e.g., a 3' tail, from nuclease degradation; 3) increase homologous recognition between the DNA sequence and the selected site for integration; and 4) increase homologous pairing between the DNA sequence and the selected site for  
10 integration.

Rad52 can be obtained in several ways including isolation of Rad52 or expression of a sequence encoding by genetic engineering methods. For example, Van Dyke et al. (1999) *Nature* 398:728, describe production and purification of hRad52 from Sf9 cells. The nucleotide sequences of Rad52 of various species are known. See, e.g., Shen et al. (1995) *Genomics* 25(1):199-206 (murine and human Rad52); Muris et al. (1994) *Mutat. Res.* 315(3):295-305 (murine and human Rad52); Park et al. (1995) *J. Biol. Chem.* 270(26):15467-15470 (human Rad52).

Fragments of Rad52 can be produced in several ways, e.g., by expression of the sequence encoding Rad52 or a portion thereof or by gene activation (the preferred method),  
20 by proteolytic digestion, or by chemical synthesis. Internal or terminal fragments of Rad52 can be generated by removing one or more nucleotides from one end (for a terminal fragment) or both ends (for an internal fragment) of a nucleic acid which encodes Rad52. Expression of the mutagenized DNA produces Rad52 polypeptide fragments. Digestion with "end-nibbling" endonucleases or with various restriction enzymes can thus generate DNA's  
25 which encode an array of Rad52 fragments. DNA's which encode fragments of a Rad52 protein can also be generated by random shearing, restriction digestion or a combination of the above-discussed methods.

Rad52 fragments can also be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example,  
30 Rad52 peptides may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or divided into overlapping fragments of a desired length.

A determination of whether a Rad52 fragment is functional can be made by known techniques. For example, to determine whether a Rad52 fragment can protect against nuclease degradation, an end-labeled linearized double stranded DNA sequence, e.g., a <sup>32</sup>P-labeled linearized double stranded DNA sequence, can be incubated with a Rad52 fragment  
35 prior to introduction of a nuclease, e.g., an exonuclease or endonuclease. The amount of released label, e.g., <sup>32</sup>P, can then be determined. The amount of released label serves as an

5 indicator of the ability of the Rad52 fragment to protect against nuclease degradation. In addition, the functionality of a Rad52 fragment can be determined based on its ability to stimulate the formation of joint molecules. The functionality of a Rad52 fragment can be analyzed *in vitro* by stimulation of hRad51-driven joint molecule formation as described in Benson et al. (1998) *Nature* 391:401-404. Briefly, hRad51 is first preincubated with circular  
10 ssDNA and then <sup>32</sup>P-labeled linear duplex DNA is added. The formation of joint molecules can be determined by electrophoresis. The addition of Rad52 stimulates the formation of joint molecules as compared to joint molecule formation in the absence of Rad52. Thus, it can be determined if a Rad52 fragment is functional if it stimulates joint molecule formation as compared to joint molecule formation in its absence. Moreover, the extent of stimulation  
15 by a Rad52 fragment can be compared to the extent of full-length Rad52 stimulation. In addition, the functionality of a Rad52 fragment can be determined based on its ability to increase resistance to ionizing radiation and to increase rates of homologous recombination when overexpressed in cultured monkey cells as described in Park (1995) *J. Biol. Chem.* 270:15467-15470.

#### 20 Agents Which Inhibit Non-Homologous End Joining

An agents which inhibits non-homologous end joining can be used to provide a DNA sequence at a selected site in target DNA at a higher rate than would occur in its absence.  
25 Non-homologous end joining can lead to imprecise fusion between double stranded ends, e.g., the rejoined ends can have insertions or deletions. An agent which inhibits non-homologous end joining can be any agent which inhibits expression of and/or an activity of a molecule involved in a non-homologous end joining pathway. For example, a complex of Mre11, Rad50 and Nbs1 is involved in non-homologous end joining. Thus, for example, by  
30 inhibiting formation of this complex, e.g., by binding any of these proteins or inhibiting expression of any of these proteins, non-homologous end joining can be inhibited. In addition, other proteins involved in non-homologous end joining include Ku proteins, e.g., Ku70 or Ku80, Ligase 4 (Lig4) and Xrcc4.

## 5 Ku Inactivating Agents

Providing a Ku inactivating agent with a DNA sequence at a selected site in a target DNA, e.g., a selected site in chromosomal DNA, can provide a higher rate of alteration of the site, e.g., homologous recombination, than would occur in its absence. Ku is a heterodimer  
10 of approximately 70 kDa and 80 kDa that binds to DNA discontinuities and plays a role in double-strand break repair by non-homologous end joining. "Ku80" can also be referred to as "Ku86".

A Ku inactivating agent can inhibit Ku expression or a Ku activity. Preferably, a Ku inactivating agent interacts, e.g., binds, Ku or a nucleotide sequence encoding Ku, to inhibit  
15 Ku expression or a Ku activity. Preferably, Ku-dependent non-homologous end joining is inhibited. A Ku inhibiting agent can inhibit Ku70, Ku80 or both.

Agents which can be used to inactivate Ku include anti-Ku antibodies and Ku-binding molecules, e.g., randomly generated peptides which bind to Ku, Ku binding oligomers and polymers, and antisense Ku nucleic acid molecules. Preferably, the agent which inactivates  
20 Ku is an agent which can be administered locally such as anti-Ku antibodies and Ku-binding molecules, e.g., randomly-generated peptides which bind to Ku, and Ku binding oligomers or polymers.

Preferably, the Ku inactivating agent interacts with, e.g., binds to, Ku. Agents which interact with the Ku protein can inactivate Ku locally at the site of alteration.

For example, a Ku inactivating agent is introduced into a cell in close proximity to the  
25 DNA sequence and the targeted DNA to thereby inhibit Ku locally at the site of homologous recombination. A Ku inactivating agent can be introduced to a cell in a mixture which includes the double stranded DNA sequence, it can be introduced immediately prior to or after administration of the DNA sequence or it can be covalently linked to the DNA sequence  
30 or proteins associated with the DNA sequence, e.g., Rad52 or a fragment thereof. Cells can also be preincubated with a Ku inactivating agent such as an anti-Ku antibody or an antisense Ku nucleic acid molecule.

### Anti-Ku Antibodies

35 An anti-Ku antibody or fragment thereof can be used to bind Ku, and thereby reduce a Ku activity. Anti-Ku antibodies can be administered such that they interact with Ku locally

5 at the site of alteration but do not inhibit Ku expression generally in the cell. Anti-Ku antibodies include anti-Ku70 and anti-Ku80 antibodies.

A Ku protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind Ku using standard techniques for polyclonal and monoclonal antibody preparation. The full-length Ku protein can be used or, alternatively, antigenic  
10 peptide fragments of Ku can be used as immunogens.

Typically, Ku or a Ku peptide is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, a Ku protein obtained by expression of the sequence encoding Ku or by gene activation, or a chemically synthesized Ku peptide.

15 See, e.g., U.S. Patent No. 5,460,959; and co-pending U.S. applications USSN 08/334,797; USSN 08/231,439; USSN 08/334,455; and USSN 08/928,881 which are hereby expressly incorporated by reference in their entirety. The nucleotide and amino acid sequences of Ku are known and described, for example, in Takiguchi et al. (1996) *Genomics* 35(1):129-135. The preparation can further include an adjuvant, such as Freund's complete or incomplete  
20 adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic Ku preparation induces a polyclonal anti-Ku antibody response.

Anti-Ku antibodies or fragments thereof can be used as a Ku inactivating agent. Examples of anti-Ku antibody fragments include F(v), Fab, Fab' and F(ab')<sub>2</sub> fragments which can be generated by treating the antibody with an enzyme such as pepsin. The term  
25 "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of Ku. A monoclonal antibody composition thus typically displays a single binding affinity for a particular Ku protein with which it immunoreacts.

30 Additionally, anti-Ku antibodies produced by genetic engineering methods, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, can be used. Such chimeric and humanized monoclonal antibodies can be produced by genetic engineering using standard DNA techniques known in the art, for example using methods described in  
35 Robinson et al. International Application No. PCT/US86/02269; Akira, et al. European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al. European Patent Application 173,494; Neuberger et al. PCT International Publication No.



5 WO 86/01533; Cabilly et al. U.S. Patent No. 4,816,567; Cabilly et al. European Patent Application 125,023; Better et al., *Science* 240:1041-1043, 1988; Liu et al., *PNAS* 84:3439-3443, 1987; Liu et al., *J. Immunol.* 139:3521-3526, 1987; Sun et al. *PNAS* 84:214-218, 1987; Nishimura et al., *Canc. Res.* 47:999-1005, 1987; Wood et al., *Nature* 314:446-449, 1985; and Shaw et al., *J. Natl. Cancer Inst.* 80:1553-1559, 1988); Morrison, S. L., *Science* 229:1202-1207, 1985; Oi et al., *BioTechniques* 4:214, 1986; Winter U.S. Patent 5,225,539; Jones et al., *Nature* 321:552-525, 1986; Verhoeyan et al., *Science* 239:1534, 1988; and Beidler et al., *J. Immunol.* 141:4053-4060, 1988.

In addition, a human monoclonal antibody directed against Ku can be made using standard techniques. For example, human monoclonal antibodies can be generated in transgenic mice or in immune deficient mice engrafted with antibody-producing human cells. Methods of generating such mice are describe, for example, in Wood et al. PCT publication WO 91/00906, Kucherlapati et al. PCT publication WO 91/10741; Lonberg et al. PCT publication WO 92/03918; Kay et al. PCT publication WO 92/03917; Kay et al. PCT publication WO 93/12227; Kay et al. PCT publication 94/25585; Rajewsky et al. Pct publication WO 94/04667; Ditullio et al. PCT publication WO 95/17085; Lonberg, N. et al. (1994) *Nature* 368:856-859; Green, L.L. et al. (1994) *Nature Genet.* 7:13-21; Morrison, S.L. et al. (1994) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Bruggeman et al. (1993) *Year Immunol* 7:33-40; Choi et al. (1993) *Nature Genet.* 4:117-123; Tuailon et al. (1993) *PNAS* 90:3720-3724; Bruggeman et al. (1991) *Eur J Immunol* 21:1323-1326); Duchosal et al. PCT publication WO 93/05796; U.S. Patent Number 5,411,749; McCune et al. (1988) *Science* 241:1632-1639), Kamel-Reid et al. (1988) *Science* 242:1706; Spanopoulou (1994) *Genes & Development* 8:1030-1042; Shinkai et al. (1992) *Cell* 68:855-868). A human antibody-transgenic mouse or an immune deficient mouse engrafted with human antibody-producing cells or tissue can be immunized with Ku or an antigenic Ku peptide and splenocytes from these immunized mice can then be used to create hybridomas. Methods of hybridoma production are well known.

Human monoclonal antibodies against Ku can also be prepared by constructing a combinatorial immunoglobulin library, such as a Fab phage display library or a scFv phage display library, using immunoglobulin light chain and heavy chain cDNAs prepared from mRNA derived from lymphocytes of a subject. See, e.g., McCafferty et al. PCT publication WO 92/01047; Marks et al. (1991) *J. Mol. Biol.* 222:581-597; and Griffiths et al. (1993) *EMBO J* 12:725-734. In addition, a combinatorial library of antibody variable regions can

5 be generated by mutating a known human antibody. For example, a variable region of a human antibody known to bind Ku, can be mutated, by for example using randomly altered mutagenized oligonucleotides, to generate a library of mutated variable regions which can then be screened to bind to Ku. Methods of inducing random mutagenesis within the CDR regions of immunoglobulin heavy and/or light chains, methods of crossing randomized heavy and light chains to form pairings and screening methods can be found in, for example, Barbas  
10 et al. PCT publication WO 96/07754; Barbas et al. (1992) *Proc. Nat'l Acad. Sci. USA* 89:4457-4461.

The immunoglobulin library can be expressed by a population of display packages, preferably derived from filamentous phage, to form an antibody display library. Examples of  
15 methods and reagents particularly amenable for use in generating antibody display library can be found in, for example, Ladner et al. U.S. Patent No. 5,223,409; Kang et al. PCT publication WO 92/18619; Dower et al. PCT publication WO 91/17271; Winter et al. PCT publication WO 92/20791; Markland et al. PCT publication WO 92/15679; Breitling et al. PCT publication WO 93/01288; McCafferty et al. PCT publication WO 92/01047; Garrard et  
20 al. PCT publication WO 92/09690; Ladner et al. PCT publication WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum Antibod Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *supra*; Hawkins et al. (1992) *J Mol Biol* 226:889-896; Clackson et al. (1991) *Nature* 352:624-628; Gram et al. (1992) *PNAS* 89:3576-3580; Garrard et al. (1991) *Bio/Technology* 9:1373-1377; Hoogenboom et al.  
25 (1991) *Nuc Acid Res* 19:4133-4137; and Barbas et al. (1991) *PNAS* 88:7978-7982. Once displayed on the surface of a display package (e.g., filamentous phage), the antibody library is screened to identify and isolate packages that express an antibody that binds Ku. In a preferred embodiment, the primary screening of the library involves panning with an immobilized Ku and display packages expressing antibodies that bind immobilized Ku are  
30 selected.

Monoclonal antibodies to Ku are also commercially available from, for example, Neomarkers (Fremont, CA).

### Ku-Binding Molecules

5 Molecules which bind Ku such as Ku-binding peptides, e.g., randomly generated peptides, and Ku-binding oligomers or polymers can be used as Ku inactivating agents. Such molecules can bind to the Ku protein and thereby inhibit at least one activity of Ku such as non-homologous end-joining.

10 Examples of Ku-binding oligomers are set forth in WO 99/33971, the contents of which is incorporated herein by reference. Such oligomers can be composed of nucleotides, nucleotide analogs, or a combination. Preferably, the oligomers are composed of ribonucleotides. These Ku oligomers can be used to bind Ku or to identify proteins that interact with Ku. Methods of identifying Ku binding peptides using these oligomers are described in WO99/33971.

15 In addition, randomly generated peptides can be screened for the ability to bind Ku. For example, various techniques are known in the art for screening generated mutant gene products. Techniques for screening large gene libraries often include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the genes under conditions in which detection of a desired activity, 20 e.g., binding to Ku, facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the techniques described below is amenable to high throughput analysis for screening large numbers of sequences created, e.g., by random mutagenesis techniques.

#### 25 Display Libraries

In another approach to screening for Ku binding peptides, the candidate peptides are displayed on the surface of a cell or viral particle, and the ability of particular cells or viral particles to bind a Ku protein via the displayed product is detected in a "panning assay". For example, the gene library can be cloned into the gene for a surface membrane protein of a 30 bacterial cell, and the resulting fusion protein detected by panning (Ladner et al., WO 88/06630; Fuchs et al. (1991) *Bio/Technology* 9:1370-1371; and Goward et al. (1992) *TIBS* 18:136-140). In a similar fashion, a detectably labeled ligand can be used to score for potentially functional peptide homologs. Fluorescently labeled ligands can be used to detect homolog which retain ligand-binding activity. The use of fluorescently labeled ligands, 35 allows cells to be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, to be separated by a fluorescence-activated cell sorter.

5 A gene library can be expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at concentrations well over  $10^{13}$  phage per milliliter, a large number of phage can be screened at one time. Second, since each infectious phage  
10 displays a gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd., and f1 are most often used in phage display libraries. Either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle. Foreign  
15 epitopes can be expressed at the NH<sub>2</sub>-terminal end of pIII and phage bearing such epitopes recovered from a large excess of phage lacking this epitope (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) *J. Biol. Chem.* 267:16007-16010; Griffiths et al. (1993) *EMBO J* 12:725-734; Clackson et al. (1991) *Nature* 352:624-628; and Barbas et al. (1992) *PNAS* 89:4457-4461).

20 A common approach uses the maltose receptor of *E. coli* (the outer membrane protein, LamB) as a peptide fusion partner (Charbit et al. (1986) *EMBO* 5, 3029-3037). Oligonucleotides have been inserted into plasmids encoding the LamB gene to produce peptides fused into one of the extracellular loops of the protein. These peptides are available for binding to ligands, e.g., to antibodies, and can elicit an immune response when the cells  
25 are administered to animals. Other cell surface proteins, e.g., OmpA (Schorr et al. (1991) *Vaccines* 9, pp. 387-392), PhoE (Agterberg, et al. (1990) *Gene* 88, 37-45), and PAL (Fuchs et al. (1991) *Bio/Tech* 9, 1369-1372), as well as large bacterial surface structures have served as vehicles for peptide display. Peptides can be fused to pilin, a protein which polymerizes to form the pilus-a conduit for interbacterial exchange of genetic information (Thiry et al.  
30 (1989) *Appl. Environ. Microbiol.* 55, 984-993). Because of its role in interacting with other cells, the pilus provides a useful support for the presentation of peptides to the extracellular environment. Another large surface structure used for peptide display is the bacterial motive organ, the flagellum. Fusion of peptides to the subunit protein flagellin offers a dense array of many peptides copies on the host cells (Kuwajima et al. (1988) *Bio/Tech.* 6, 1080-1083).  
35 Surface proteins of other bacterial species have also served as peptide fusion partners. Examples include the *Staphylococcus* protein A and the outer membrane protease IgA of



5 *Neisseria* (Hansson et al. (1992) *J. Bacteriol.* 174, 4239-4245 and Klauser et al. (1990) *EMBO J.* 9, 1991-1999).

In the filamentous phage systems and the LamB system described above, the physical link between the peptide and its encoding DNA occurs by the containment of the DNA within a particle (cell or phage) that carries the peptide on its surface. Capturing the peptide  
10 captures the particle and the DNA within. An alternative scheme uses the DNA-binding protein LacI to form a link between peptide and DNA (Cull et al. (1992) *PNAS USA* 89:1865-1869). This system uses a plasmid containing the LacI gene with an oligonucleotide cloning site at its 3'-end. Under the controlled induction by arabinose, a LacI-peptide fusion protein is produced. This fusion retains the natural ability of LacI to bind to a short DNA sequence  
15 known as LacO operator (LacO). By installing two copies of LacO on the expression plasmid, the LacI-peptide fusion binds tightly to the plasmid that encoded it. Because the plasmids in each cell contain only a single oligonucleotide sequence and each cell expresses only a single peptide sequence, the peptides become specifically and stably associated with the DNA sequence that directed its synthesis. The cells of the library are gently lysed and the  
20 peptide-DNA complexes are exposed to a matrix of immobilized receptor to recover the complexes containing active peptides. The associated plasmid DNA is then reintroduced into cells for amplification and DNA sequencing to determine the identity of the peptide ligands. As a demonstration of the practical utility of the method, a large random library of dodecapeptides was made and selected on a monoclonal antibody raised against the opioid  
25 peptide dynorphin B. A cohort of peptides was recovered, all related by a consensus sequence corresponding to a six-residue portion of dynorphin B. (Cull et al. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89-1869)

This scheme, sometimes referred to as peptides-on-plasmids, differs in two important ways from the phage display methods. First, the peptides are attached to the C-terminus of  
30 the fusion protein, resulting in the display of the library members as peptides having free carboxy termini. Both of the filamentous phage coat proteins, pIII and pVIII, are anchored to the phage through their C-termini, and the guest peptides are placed into the outward-extending N-terminal domains. In some designs, the phage-displayed peptides are presented right at the amino terminus of the fusion protein. (Cwirla, et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 6378-6382) A second difference is the set of biological biases affecting the  
35 population of peptides actually present in the libraries. The LacI fusion molecules are confined to the cytoplasm of the host cells. The phage coat fusions are exposed briefly to the

5 cytoplasm during translation but are rapidly secreted through the inner membrane into the periplasmic compartment, remaining anchored in the membrane by their C-terminal hydrophobic domains, with the N-termini, containing the peptides, protruding into the periplasm while awaiting assembly into phage particles. The peptides in the LacI and phage libraries may differ significantly as a result of their exposure to different proteolytic  
10 activities. The phage coat proteins require transport across the inner membrane and signal peptidase processing as a prelude to incorporation into phage. Certain peptides exert a deleterious effect on these processes and are underrepresented in the libraries (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251). These particular biases are not a factor in the LacI display system.

15 The number of small peptides available in recombinant random libraries is enormous. Libraries of  $10^7$ - $10^9$  independent clones are routinely prepared. Libraries as large as  $10^{11}$  recombinants have been created, but this size approaches the practical limit for clone libraries. This limitation in library size occurs at the step of transforming the DNA containing randomized segments into the host bacterial cells. To circumvent this limitation,  
20 an *in vitro* system based on the display of nascent peptides in polysome complexes has recently been developed. This display library method has the potential of producing libraries 3-6 orders of magnitude larger than the currently available phage/phagemid or plasmid libraries. Furthermore, the construction of the libraries, expression of the peptides, and screening, is done in an entirely cell-free format.

25 In one application of this method (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251), a molecular DNA library encoding  $10^{12}$  decapeptides was constructed and the library expressed in an *E. coli* S30 *in vitro* coupled transcription/translation system. Conditions were chosen to stall the ribosomes on the mRNA, causing the accumulation of a substantial proportion of the RNA in polysomes and yielding complexes containing nascent peptides still  
30 linked to their encoding RNA. The polysomes are sufficiently robust to be affinity purified on immobilized receptors in much the same way as the more conventional recombinant peptide display libraries are screened. RNA from the bound complexes is recovered, converted to cDNA, and amplified by PCR to produce a template for the next round of synthesis and screening. The polysome display method can be coupled to the phage display  
35 system. Following several rounds of screening, cDNA from the enriched pool of polysomes was cloned into a phagemid vector. This vector serves as both a peptide expression vector, displaying peptides fused to the coat proteins, and as a DNA sequencing vector for peptide

5 identification. By expressing the polysome-derived peptides on phage, one can either continue the affinity selection procedure in this format or assay the peptides on individual clones for binding activity in a phage ELISA, or for binding specificity in a completion phage ELISA (Barret, et al. (1992) *Anal. Biochem* 204,357-364). To identify the sequences of the active peptides one sequences the DNA produced by the phagemid host.

#### Antisense Ku Nucleic Acid Sequences

10 Nucleic acid molecules which are antisense to a nucleotide encoding Ku can be used as an inactivating agent which inhibits Ku expression. An "antisense" nucleic acid includes a nucleotide sequence which is complementary to a "sense" nucleic acid encoding Ku, e.g.,  
15 complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can form hydrogen bonds with a sense nucleic acid. The antisense nucleic acid can be complementary to an entire Ku coding strand, or to only a portion thereof. For example, an antisense nucleic acid molecule which antisense to the "coding region" of the coding strand of a nucleotide sequence  
20 encoding Ku can be used.

Given the coding strand sequences encoding Ku disclosed in, for example, Takiguchi et al. (1996) *Genomics* 35(1):129-135 and Genbank Accession Number L35932, antisense nucleic acids can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of Ku  
25 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of Ku mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of Ku mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid can be constructed using chemical synthesis  
30 and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and  
35 acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-

5 (carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-  
 10 mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense  
 15 nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest.

### Exogenous DNA Sequences

20

The DNA sequence to be provided, e.g., introduced, into the cell can alter a target sequence in a cell. For example, a selected DNA sequence can be introduced which differs from the target DNA by less than 10, 8, 5, 4, 3, 2, or by a single nucleotide, e.g., by a substitution, a deletion or an insertion. The selected DNA sequence can also differ from the  
 25 target sequence by more than one nucleotide, e.g., differs from the target sequence by a number of nucleotides such that the selected DNA sequence has an unpaired region, e.g., a loop out region. These alterations can modify target sequence expression. Modified sequence expression includes: activating a sequence, e.g., a coding DNA sequence, e.g., a coding sequence normally found in a cell, which is normally silent (unexpressed) in the cell;  
 30 increasing expression of a sequence, e.g., a coding DNA sequence, e.g., a coding sequence normally found in a cell, which is expressed at lower than normal levels in the cell; expressing a sequence, e.g., a coding DNA sequence, e.g., a coding sequence normally found in a cell, which is normally expressed in a defective form in the cell; changing the pattern of regulation or induction of a sequence, e.g., a coding DNA sequence, e.g., a coding sequence  
 35 normally found in a cell, such that it is different than the cell's normal pattern; reducing expression of a sequence, e.g., a coding DNA sequence, e.g., a coding sequence normally found in a cell, from normal expression levels in the cell.



5 A selected DNA sequence can be introduced which differs from the target DNA by less than 10, 8, 5, 4, 3, 2, or by a single nucleotide, e.g., by a substitution, a deletion or an insertion. For example, the targeted sequence can differ from the wild-type sequence by less than 10, 8, 5, 4, 3, 2, or by a single nucleotide. Preferably, the targeted sequence differs from the wild-type sequence by a point mutation, e.g., a mutation arising from an insertion,  
10 deletion or substitution. Preferably, the mutation in the target sequence, e.g., a gene, is associated with, e.g., controls, a disease or a dysfunction. Examples of genes in which a mutation, e.g., a point mutation, has been associated with a disease or dysfunction include, but are not limited to, cystic fibrosis transmembrane regulator (CFTR) gene,  $\beta$ -globin gene, Factor VIII gene, Factor IX gene, von Willebrand factor gene, xeroderma pigmentosum group G (XP-G) gene. The selected DNA sequence for altering the target sequence can  
15 include a normal wild-type sequence which can correct the mutation. There are several genetic disorders and genes which can be altered according to the methods described herein.

In another aspect, the selected DNA sequence can also differ from the target sequence by more than one nucleotide, e.g., differs from the target sequence by a number of  
20 nucleotides such that the selected DNA sequence has an unpaired region, e.g., a loop out region. For example, the selected DNA sequence can be homologously recombined with a preselected element of the target, e.g., if one is a regulatory element and the other is a sequence which encodes a protein, the regulatory element controls expression of the protein encoding sequence. The selected DNA sequence can be a regulatory sequence, e.g., an  
25 exogenous regulatory sequence. Regulatory sequences include a promoter, an enhancer, an UAS, a scaffold-attachment region and a transcription binding site. In addition, the selected DNA sequence can also include an exon, an intron, a CAP site, a nucleotide sequence ATG, a marker, e.g., a selection marker, a splice-donor site and/or encoding DNA in frame with the target sequence. The selected DNA sequence can also include a coding region, e.g., DNA  
30 sequence encoding a protein.

The coding sequence can be endogenous, e.g., the selected DNA sequence is a regulatory sequence, or the selected DNA sequence can include the coding region, i.e., the coding region is exogenous. The coding region can encode various proteins. Examples of such proteins include: erythropoietin, calcitonin, growth hormone, insulin, insulinotropin,  
35 insulin-like growth factors, parathyroid hormone,  $\alpha$ 2-interferon (IFNA2),  $\beta$ -interferon,  $\gamma$ -interferon, nerve growth factors, FSH $\beta$ , TGF- $\beta$ , tumor necrosis factor, glucagon, bone growth factor-2, bone growth factor-7, TSH- $\beta$ , interleukin 1, interleukin 2, interleukin 3, interleukin

5 6, interleukin 11, interleukin 12, CSF-granulocyte (GCSF), CSF-macrophage, CSF-granulocyte/macrophage, immunoglobulins, catalytic antibodies, protein kinase C, glucocerebrosidase, superoxide dismutase, tissue plasminogen activator, urokinase, antithrombin III, DNase,  $\alpha$ -galactosidase, tyrosine hydroxylase, blood clotting factor V, blood clotting factor VII, blood clotting factor VIII, blood clotting factor IX, blood clotting  
10 factor X, blood clotting factor XIII, apolipoprotein E, apolipoprotein A-I, globins, low density lipoprotein receptor, IL-2 receptor, IL-2 antagonists,  $\alpha$ -1-antitrypsin, immune response modifiers,  $\beta$ -glucocereamidase,  $\alpha$ -iduronidase,  $\alpha$ -L-iduronidase, glucosamine-N-sulfatase,  $\alpha$ -N-acetylglucosaminidase, acetylcoenzymeA: $\alpha$ -glucosamine-N-acetyltransferase, N-acetylglucosamine-6-sulfatase,  $\beta$ -galactosidase,  $\beta$ -glucuronidase, N-acetylgalactosamine-  
15 6-sulfatase, and soluble CD4. Sequences encoding these proteins are known.

The term exogenous refers to a sequence which is introduced into a cell by the methods described herein. The exogenous sequence can have a sequence identical or different from an endogenous sequence present in the cell.

Preferably, the DNA sequence is a linear sequence.

#### 20 Targeting Sequence or Sequences

Targeting sequence or sequences are DNA sequences which permit homologous recombination into the genome of a cell containing the targeted sequence, e.g., the targeted gene. The term "targeting sequence" and "flanking sequence" are used interchangeably  
25 herein. Targeting sequences are, generally, DNA sequences which are homologous to (i.e., identical or sufficiently similar to cellular DNA such that the targeting sequence and cellular DNA can undergo homologous recombination) DNA sequences normally present in the genome of a cells as obtained. For example, the targeting sequence can be sufficiently homologous to: coding or noncoding DNA, a sequence lying upstream of the transcriptional  
30 start site, within, or downstream of the transcriptional stop site of a gene of interest, or sequences present in the genome through a previous modification. The targeting sequence or sequences used are selected with reference to the site into which the selected DNA sequence is to be inserted or the site into which the targeted sequence is to be altered.

One or more targeting sequences can be employed. Preferably, the selected DNA  
35 sequence is flanked by two targeting sequences. A targeting sequence can be within a gene or coding sequence (such as, the sequences of an exon and/or intron), immediately adjacent to a coding sequence of a gene (e.g., with less than 10, 5, 4, 3, 2, 1 or no additional nucleotides

5 between the targeting sequence and the coding region of the gene), upstream of a coding  
sequence of a gene (such as the sequences of the upstream non-coding region or endogenous  
promoter sequences), or upstream of and at a distance from the coding sequence of a gene  
(such as, sequences upstream of the endogenous promoter). The targeting sequence or  
sequences can include those regions of the targeted sequence presently known or sequenced  
10 and/or regions further upstream which are structurally uncharacterized but can be mapped  
using restriction enzymes and determined by one skilled in the art.

A targeting sequence can be used to insert a DNA sequence which includes a  
regulatory sequence immediately adjacent to, upstream, or at a substantial distance from the  
coding sequence of an endogenous gene. Alternatively or additionally, sequences which  
15 affect the structure or stability of the RNA or protein produced can be replaced, removed,  
added, or otherwise modified by targeting. For example, RNA stability elements, splice sites,  
and/or leader sequences of RNA molecules can be modified to improve or alter the function,  
stability, and/or translatability of an RNA molecule. Protein sequences may also be altered,  
such as signal sequences, propeptide sequences, active sites, and/or structural sequences for  
20 enhancing or modifying transport, secretion, or functional properties of a protein. A protein  
sequence can also be altered, e.g., corrected, by targeting a site in the gene encoding the  
protein which includes a mutation, e.g., a point mutation.

In one aspect, the targeting sequence can be homologous to a portion of human  
follicle stimulating hormone  $\beta$  (FSH $\beta$ ). FSH is a gonadotrophin which plays an essential role  
25 in the maintenance and development of oocytes and spermatozoa in normal reproductive  
physiology. FSH includes two subunits,  $\alpha$  and  $\beta$ , the latter being responsible for FSH's  
biological specificity. The target site to which a given targeting sequence is homologous can  
reside within an exon and/or intron of the FSH $\beta$  gene, upstream of and immediately adjacent  
to the FSH $\beta$ -coding region, or upstream of and at a distance from the FSH $\beta$ -coding region.  
30 For example, the first of the two targeting sequences (or the entire targeting sequence, if there  
is only one targeting sequence in the construct) can be derived from the genomic regions  
upstream of the FSH $\beta$ -coding sequences. For example, this targeting sequence can include a  
portion of SEQ ID NO:1, e.g., at least 20, 30, 50, 100, or 1000 consecutive nucleotides from  
the sequence corresponding to positions -7,454 to -1,417 (SEQ ID NO:2) or to positions -696  
35 to -155 (SEQ ID NO:3). The second of the two targeting sequences can target a genomic  
region upstream of the coding sequence (e.g., also contain a portion of SEQ ID NO:2 or 3), or  
target an exon or intron of the gene. Sequences which can be used to target FSH $\beta$  are further

5 described in U.S. Serial Number 09/305,639, the contents of which is incorporated herein in its entirety.

The targeting sequence can be homologous to a portion of human interferon- $\alpha$ 2 (IFN $\alpha$ 2). Interferon- $\alpha$  constitutes a complex gene family with 14 genes clustered on the short arm of chromosome 9. None of these genes, including IFN $\alpha$ 2 gene, have introns. Interferon- $\alpha$  is produced by macrophages, T cells and B cells as wells as many other cell types. Interferon- $\alpha$  has considerable anti-viral effects, and has been shown to be efficacious in treating infections by papilloma virus, hepatitis B and C viruses, vaccina, herpes simplex virus, herpes zoster varicellosus virus and rhinovirus.

15 The target site to which a given targeting sequence is homologous can reside within the coding region of the IFN $\alpha$ 2 gene, upstream of and immediately adjacent to the coding region, or upstream and at a distance from the coding region. For example, the first of the two targeting sequences (or the entire targeting sequence, if there is only one targeting sequence in the construct) can be derived from the genomic regions upstream of the IFN $\alpha$ 2-coding sequences. For example, this targeting sequence can include a portion (e.g., at least 20, 50, 100 or 1000 consecutive nucleotides) of SEQ ID NO:4, which corresponds to nucleotides -4074 to -511 of the IFN $\alpha$ 2 gene. The second of the two targeting sequences may target a genomic region upstream of the coding sequence itself. By way of example, the second targeting sequence may contain at its 3' end, an exogenous coding region identical to the first few codons of the IFN $\alpha$ 2 coding sequence. Upon homologous recombination, the exogenous coding region recombines with the targeted part of the endogenous IFN $\alpha$ 2 coding sequence. Sequences which can be used to target IFN $\alpha$ 2 are further described in U.S. Serial Number 09/305,638, the contents of which is incorporated herein in its entirety.

30 In another aspect, the targeting sequence can be homologous to a portion of human granulocyte colony-stimulating factor (GCSF). GCSF is a cytokine that stimulates the proliferation and differentiation of hematopoietic progenitor cells committed to the neutrophil/granulocyte lineage. GCSF is routinely used in the prevention of chemotherapy-induced neutropenia and in association with bone marrow transplantation. Chronic idiopathic and congenital neutropenic disorders also show improvement after GCSF injection. The target site to which a given targeting sequence is homologous can reside within an exon and/or intron of the GCSF gene, upstream of and immediately adjacent to the GCSF coding region, or upstream of and at a distance from the GCSF coding region.



5 For example, the first of the two targeting sequences in the construct (or  
the entire targeting sequence, if there is only one targeting sequence in the construct) can be  
homologous to the genomic regions upstream of the GCSF-coding sequences. For example,  
this targeting sequence can contain a portion of SEQ ID NO:5, which corresponds to  
nucleotides -6,578 to 101 of human GCSF gene (e.g., at least 20, 50, 100, or 1000  
10 consecutive nucleotides from the sequence corresponding to positions -6,578 to -364 (SEQ  
ID NO:6)). The second of the two targeting sequences in the construct may target a genomic  
region upstream of the coding sequence (e.g., also contain a portion of SEQ ID NO:6), or  
target an exon or intron of the gene. Sequences which can be used to target GCSF are further  
described in U.S. Serial Number 09/305,384, the contents of which is incorporated herein in  
15 its entirety.

## Regulatory Sequence

A DNA sequence can include a regulatory sequence. The regulatory sequence can  
include one or more promoters (such as a constitutive or inducible promoter), enhancers, an  
20 UAS, scaffold-attachment regions or matrix attachment sites, negative regulatory elements,  
transcription factor binding sites, or combinations of these sequences.

The regulatory sequence can contain an inducible promoter such that cells as  
produced or as introduced into an individual can be induced to express a product, e.g., the cell  
does not express the product but can be induced to express it. The regulatory sequence can  
25 contain an inducible promoter such that the product is expressed upon introduction of the  
regulatory sequence. The regulatory sequence can be a cellular or viral sequence. Such  
regulatory sequences include, but are not limited to, those that regulate the expression of  
SV40 early or late genes, adenovirus major late genes, the mouse metallothionein-I gene, the  
elongation factor-1 $\alpha$  gene, cytomegalovirus genes, collagen genes, actin genes,  
30 immunoglobulin genes,  $\gamma$  actin gene, transcription activator YY1 gene, fibronectin gene, or  
the HMG-CoA reductase gene. The regulatory sequence can further contain a transcription  
factor binding site, such as a TATA Box, CCAAT Box, AP1, Sp1 or NF- $\kappa$ B binding site.

## Additional DNA Sequence Elements

35 The DNA sequence can further include one or more exons. An exon is a DNA  
sequence which is copied into RNA and is present in a mature mRNA molecule. An exons

5 can contain DNA which encodes one or more amino acids and/or partially encodes an amino acid (i.e., one or two bases of a codon). Alternatively, an exon contains DNA which corresponds to a non-coding region, e.g., a 5' non-coding region. Where the exogenous exon or exons encode one or more amino acids and/or a portion of an amino acid, the DNA sequence can be designed such that, upon transcription and splicing, the reading frame is in-  
10 frame with the second exon or coding region of a targeted gene. As used herein, in-frame means that the encoding sequences of a first exon and a second exon, when fused, join together nucleotides in a manner that does not change the appropriate reading frame of the portion of the mRNA derived from the second exon.

If the first exon of the targeted gene contains the sequence ATG to initiate translation,  
15 the exogenous exon preferably contains an ATG. In addition, an exogenous exon containing an ATG can further include one or more nucleotides such that the resulting coding region of the mRNA including the second and subsequent exons of the targeted gene is in-frame. Examples of such targeted genes in which the first exon contains an ATG include the genes encoding human erythropoietin, human growth hormone, human colony stimulating factor-  
20 granulocyte/macrophage (hGM-CSF), and human colony stimulating factor-granulocyte (hG-CSF).

A splice-donor site is a sequence which directs the splicing of one exon to another exon. Typically, a first exon lies 5' of a second exon, and a splice-donor site overlapping and flanking the first exon on its 3' side recognizes a splice-acceptor site flanking the second exon  
25 on the 5' side of the second exon. A splice-donor site can have a characteristic consensus sequence represented as: (A/C)AG GURAGU (where R denotes a purine nucleotide) with the GU in the fourth and fifth positions, being required (Jackson 1991) *Nucleic Acids Res.* 19: 3715-3798). The first three bases of the splice-donor consensus site are the last three bases of the exon. Splice-donor sites can be functionally defined by their ability to effect the  
30 appropriate reaction within the mRNA splicing pathway.

An unpaired splice-donor site is a splice-donor site which is present in a targeted sequence and is not accompanied in the DNA sequence by a splice-acceptor site positioned 3' to the unpaired splice-donor site. The unpaired splice-donor site can result in splicing to an endogenous splice-acceptor site.

35 A splice-acceptor site in a sequence which, like a splice-donor site, directs the splicing of one exon to another exon. Acting in conjunction with a splice-donor site, the splicing apparatus uses a splice-acceptor site to effect the removal of an intron. Splice-

5 acceptor sites can have a characteristic sequence represented as: YYYYYYYYYYNYAG, where Y denotes any pyrimidine and N denotes any nucleotide (Jackson (1991) *Nucleic Acids Res.* 19:3715-3798).

10 An intron is defined as a sequence of one or more nucleotides lying between two exons and which is removed, by splicing, from a precursor RNA molecule in the formation of an mRNA molecule.

The regulatory sequence can be linked to an ATG start codon, for initiating translation. Optionally, a CAP site (a specific mRNA initiation site which is associated with and utilized by the regulatory region) can be linked to the regulatory sequence and the ATG start codon. Alternatively, the CAP site associated with and utilized by the regulatory  
15 sequence is not included in the target sequence, and the transcriptional apparatus provides a new CAP site. A CAP site can usually be found approximately 25 nucleotides 3' of the TATA box. A splice-donor site can be placed immediately adjacent to an ATG, e.g., where the presence of one or more nucleotides is not required for the exogenous exon to be in-frame with the second exon of the targeted gene. DNA encoding one or more amino acids or  
20 portions of an amino acid in-frame with the coding sequence of the targeted gene, can be placed immediately adjacent to the ATG on its 3' side. As such, the splice-donor site can be placed immediately adjacent to the encoding DNA on its 3' side.

An encoding portion of a DNA sequence (e.g., in exon 1 of the DNA sequence) can encode one or more amino acids, and/or a portion of an amino acid, which are the same as  
25 those of the endogenous protein. For example, the encoding DNA sequence can correspond to the first exon of the gene of interest. The encoding DNA can alternatively encode one or more amino acids or a portion of an amino acid different from the first exon of the protein of interest, for example, where the amino acids of the first exon of the protein of interest are not critical to the activity or activities of the protein. For example, when fusions to an  
30 endogenous human erythropoietin (EPO) gene are constructed, sequences encoding the first exon of human growth hormone (hGH) can be employed. In this example, fusion of hGH exon 1 to EPO exon 2 results in the formation of a hybrid signal peptide which is functional. However, any exon of human or non-human origin in which the encoded amino acids do not prevent the function of the hybrid signal peptide can be used.

35 Where the desired product is a fusion protein of the endogenous protein and encoding sequences in the DNA sequence, the exogenous encoding DNA incorporated into the cells can include DNA which encodes one or more exons or a sequence of cDNA corresponding to

5 a translation or transcription product which is to be fused to the product of the endogenous targeted gene. Thus, targeting can be used to prepare chimeric or multifunctional proteins which combine structural, enzymatic, or ligand or receptor binding properties from two or more proteins into one polypeptide. For example, the exogenous DNA sequence can encode, e.g., an anchor to the membrane for the targeted protein or a signal peptide to provide or  
10 improve cellular secretion, leader sequences, enzymatic regions, transmembrane domain regions, co-factor binding regions or other functional regions. Examples of proteins which are not normally secreted, but which could be fused to a signal protein to provide secretion include dopa-decarboxylase, transcriptional regulatory proteins and tyrosine hydroxylase.

The DNA sequence can be obtained from sources in which it occurs in nature or can  
15 be produced, using genetic engineering techniques or synthetic processes.

#### Target Sequence

The DNA sequence, when transfected into cells, such as primary, secondary or  
20 immortalized cells, can control the expression of a desired product for example, the active or, functional portion of the protein or RNA. The DNA sequence can also encode a desired product. The product can be, for example, a hormone, a cytokine, an antigen, an antibody, an enzyme, a clotting factor, a transport protein, a receptor, a regulatory protein, a structural protein, a transcription factor, an anti-sense RNA, or a ribozyme. Additionally, the product  
25 can be a protein or a nucleic acid which does not occur in nature (i.e., a fusion protein or nucleic acid).

Such products include erythropoietin, calcitonin, growth hormone, insulin, insulinotropin, insulin-like growth factors, parathyroid hormone, interferon  $\beta$ , and interferon  $\gamma$ , nerve growth factors, FSH $\beta$ , TGF- $\beta$ , tumor necrosis factor, glucagon, bone growth factor-  
30 2, bone growth factor-7, TSH- $\beta$ , interleukin 1, interleukin 2, interleukin 3, interleukin 6, interleukin 11, interleukin 12, CSF-granulocyte, CSF-macrophage, CSF-granulocyte/macrophage, immunoglobulins, catalytic antibodies, protein kinase C, glucocerebrosidase, superoxide dismutase, tissue plasminogen activator, urokinase, antithrombin III, DNase,  $\alpha$ -galactosidase, tyrosine hydroxylase, blood clotting factors V,  
35 blood clotting factor VII, blood clotting factor VIII, blood clotting factor IX, blood clotting factor X, blood clotting factor XIII, apolipoprotein E or apolipoprotein A-I, globins, low density lipoprotein receptor, IL-2 receptor, IL-2 antagonists, alpha-1 anti-trypsin, immune



5 response modifiers,  $\beta$ -glucoceromidase,  $\alpha$ -iduronidase,  $\alpha$ L-iduronidase, glucosamine-N-sulfatase,  $\alpha$ -N-acetylglucosaminidase, acetylcoenzymeA: $\alpha$ -glucosamide-N-acetyltransferase, N-acetylglucosamine-6-sulfatase,  $\beta$ -galactosidase,  $\beta$ -glucuronidase, N-acetylgalactosamine-6-sulfatase, and soluble CD4.

#### 10 Selectable Markers and Amplification

The identification of the targeting event can be facilitated by the use of one or more selectable marker genes. These markers can be included in the DNA sequence or can be present on a different construct. Selectable markers can be divided into two categories: positively selectable and negatively selectable (in other words, markers for either positive  
15 selection or negative selection). In positive selection, cells expressing the positively selectable marker are capable of surviving treatment with a selective agent (such as neo, xanthine-guanine phosphoribosyl transferase (gpt), dhfr, adenine deaminase (ada), puromycin (pac), hygromycin (hyg), CAD which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydro-orotase glutamine synthetase (GS), multidrug resistance 1  
20 (mdr1) and histidine D (hisD), allowing for the selection of cells in which the targeting construct integrated into the host cell genome. In negative selection, cells expressing the negatively selectable marker are destroyed in the presence of the selective agent. The identification of the targeting event can be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is  
25 linked to the exogenous DNA sequence, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker (Mansour et al. (1988) *Nature* 336:348-352). Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene, the bacterial gpt  
30 gene, diphteria toxin and antisense RNA or ribozyme for mRNA that codes for a gene essential for cell survival.

A variety of selectable markers can be incorporated into primary, secondary or immortalized cells. For example, a selectable marker which confers a selectable phenotype such as drug resistance, nutritional auxotrophy, resistance to a cytotoxic agent or expression  
35 of a surface protein, can be used. Selectable marker genes which can be used include neo, gpt, dhfr, ada, pac, hyg, CAD, GS, mdr1 and hisD. The selectable phenotype conferred makes it possible to identify and isolate recipient cells.

5 Genes encoding selectable markers (e.g., *ada*, *GS*, *dhfr* and the multifunctional *CAD* gene) have the added characteristic that they enable the selection of cells containing increased copies of the selectable marker and flanking the genomic sequence. This feature provides a mechanism for significantly increasing the copy number of an adjacent or linked gene for which increased copies is desirable. Mutated versions of these sequences showing improved  
10 selection properties and other sequences leading to increased copies can also be used.

The order and number of components in the DNA sequence can vary. For example, the order can be: a first targeting sequence--selectable marker--regulatory sequence--an exon--a splice-donor site--a second targeting sequence or, in the alternative, a first targeting sequence--regulatory sequence--an exon--a splice-donor site--DNA encoding a selectable  
15 marker--a second targeting sequence. Cells that stably integrate the construct will survive treatment with the selective agent; a subset of the stably transfected cells will be homologously recombinant cells. The homologously recombinant cells can be identified by a variety of techniques, including PCR, Southern hybridization and phenotypic screening. The order of the construct can be: a first targeting sequence--selectable marker--regulatory  
20 sequence--an exon--a splice-donor site--an intron--a splice-acceptor site--a second targeting sequence.

Alternatively, the order of components in the DNA sequence can be, for example: a first targeting sequence --selectable marker 1--regulatory sequence--an exon--a splice-donor site--a second targeting sequence--selectable marker 2, or, alternatively, a first targeting  
25 sequence--regulatory sequence--an exon--a splice-donor site--selectable marker 1--a second targeting sequence--selectable marker 2. In this arrangement, selectable marker 2 can display the property of negative selection. That is, the gene product of selectable marker 2 can be selected against by growth in an appropriate media formulation containing an agent (typically a drug or metabolite analog) which kills cells expressing selectable marker 2. Recombination  
30 between the targeting sequences flanking selectable marker 1 with homologous sequences in the host cell genome results in the targeted integration of selectable marker 1, while selectable marker 2 is not integrated. Such recombination events generate cells which are stably transfected with selectable marker 1 but not stably transfected with selectable marker 2, and such cells can be selected for by growth in the media containing the selective agent  
35 which selects for selectable marker 1 and the selective agent which selects against selectable marker 2.

5           The DNA sequence also can include a positively selectable marker that allows for the selection of cells containing increased copies of that marker. The increased copies of such a marker result in the co-amplification of flanking DNA sequences. For example, the order of components can be: a first targeting sequence--a positively selectable marker which increases the number of copies--a second selectable marker (optional)--regulatory sequence--an exon--  
10 a splice-donor site--a second targeting DNA sequence. The activated gene can be further increased by the inclusion of a selectable marker gene which has the property that cells containing increased copies of the selectable marker gene can be selected for by culturing the cells in the presence of the appropriate selectable agent. The activated endogenous gene will be increases in tandem with the selectable marker gene. Cells containing many copies of the  
15 activated endogenous gene may produce very high levels of the desired protein and are useful for *in vitro* protein production and gene therapy.

          The selectable and other marker genes do not have to lie immediately adjacent to each other.

20   DNA Sequence/Homologous Recombination Enhancing Agent/Non-Homologous End Joining Inhibiting Agent Complexes

          Homologous recombination between a double stranded DNA sequence and a selected target DNA, e.g., chromosomal DNA in a cell, can be promoted by providing an agent which  
25 enhances homologous recombination, e.g., a Rad52 protein, and an agent which inhibits non-homologous end joining, e.g., a Ku inactivating agent (e.g., a anti-Ku antibody), in sufficiently close proximity to the DNA sequence and the targeted site. "Sufficiently close proximity" as used herein refers to introduction of a homologous recombination enhancing agent or an agent which inhibits non-homologous end joining or both such that the  
30 concentration of the homologous recombination enhancing agent and/or agent which inhibits non-homologous end joining is sufficient to provide a higher rate of an alteration at a targeted site, e.g., homologous recombination between a DNA sequence and a target sequence. Several methods can be used to provide the introduction of the DNA sequence, homologous recombination enhancing agent, and an agent which inhibits non-homologous end joining  
35 within close proximity of each other. By administering these compounds in close proximity of each other and the target DNA, the activity of compounds such as Rad52 and Ku inactivating molecules, e.g., an anti-Ku antibody, are localized at the site of homologous

5 recombination. For example, local inhibition of Ku activity may be preferable over whole cell inhibition of Ku activities.

The close proximity of the DNA sequence, a homologous recombination enhancing agent, and an agent which inhibits non-homologous end joining can be maintained by introducing these elements as part of a complex. For example, DNA-protein complexes can be used. The core of a DNA-protein complex can be composed of the double stranded DNA sequence which is to be introduced into the selected site in the target DNA. A homologous recombination enhancing agent, e.g., a Rad52 protein or fragment thereof, can be adhered, e.g., coated, on the DNA sequence, e.g., on the entire sequence or just the ends of the DNA sequence, e.g., on at least a portion of a single stranded protruding end of the DNA sequence. The DNA-protein complex can further include an agent which inhibits non-homologous end joining, e.g., a Ku inactivating agent such as an anti-Ku antibody, which is covalently linked to either the DNA sequence or to the homologous recombination enhancing agent. The agent which inhibits non-homologous end joining can also be non-covalently linked to the DNA sequence or to the homologous recombination enhancing agent.

20 The compounds can also be maintained in close proximity to one another by providing the DNA sequence, the homologous recombination enhancing agent and the agent which inhibits non-homologous end joining in a liposome or vesicle. For example, liposomal suspensions can also be used as pharmaceutically acceptable carriers of these elements. Liposomal suspensions can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

The DNA sequence, the homologous recombination enhancing agent and the agent which inhibits non-homologous end joining can also be part of a mixed solution which can be microinjected into a cell or each of these compounds can be introduced in quick succession to the others such that all three of these compounds are present in the cell at the same time. Other methods of introducing one or more of these compounds include receptor-mediated delivery, electroporation and calcium phosphate precipitation.

### Cells

Primary and secondary cells to be transfected can be obtained from a variety of tissues and include cell types which can be maintained and propagated in culture. For example, primary and secondary cells which can be transfected include fibroblasts, keratinocytes, epithelial cells (e.g., mammary epithelial cells, intestinal epithelial cells), endothelial cells,



5 glial cells, neural cells, formed elements of the blood (e.g., lymphocytes, bone marrow cells), muscle cells and precursors of these somatic cell types. Primary cells are preferably obtained from the individual to whom the transfected primary or secondary cells are administered (i.e., an autologous cell). However, primary cells may be obtained from a donor (other than the recipient) of the same species (i.e., an allogeneic cell) or another species (i.e., a xenogeneic  
10 cell) (e.g., mouse, rat, rabbit, cat, dog, pig, cow, bird, sheep, goat, horse).

Primary or secondary cells of vertebrate, particularly mammalian, origin can be transfected with an exogenous DNA sequence, e.g., an exogenous DNA sequence encoding a therapeutic protein, and produce an encoded therapeutic protein stably and reproducibly, both *in vitro* and *in vivo*, over extended periods of time. In addition, the transfected primary and  
15 secondary cells can express the encoded product *in vivo* at physiologically relevant levels, cells can be recovered after implantation and, upon reculturing, to grow and display their preimplantation properties.

Alternatively, primary or secondary cells of vertebrate, particularly mammalian, origin can be transfected with an exogenous DNA sequence which includes a regulatory  
20 sequence. Examples of such regulatory sequences include one or more of: a promoter, an enhancer, an UAS, a scaffold attachment region or a transcription binding site. The targeting event can result in the insertion of the regulatory sequence of the DNA sequence, placing a targeted endogenous gene under their control (for example, by insertion of either a promoter or an enhancer, or both, upstream of the endogenous gene or regulatory region). Optionally,  
25 the targeting event can simultaneously result in the deletion of an endogenous regulatory sequence, such as the deletion of a tissue-specific negative regulatory sequence, of a gene. The targeting event can replace an existing regulatory sequence; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally-occurring elements, or displays a pattern of regulation or  
30 induction that is different from the corresponding nontransfected cell. In this regard, the naturally occurring sequences are deleted and new sequences are added. Alternatively, the endogenous regulatory sequences are not removed or replaced but are disrupted or disabled by the targeting event, such as by targeting the exogenous sequences within the endogenous regulatory elements. Introduction of a regulatory sequence by homologous recombination  
35 can result in primary or secondary cells expressing a therapeutic protein which it does not normally express. In addition, targeted introduction of a regulatory sequence can be used for cells which make or contain the therapeutic protein but in lower quantities than normal (in

5 quantities less than the physiologically normal lower level) or in defective form, and for cells which make the therapeutic protein at physiologically normal levels, but are to be augmented or enhanced in their content or production.

The transfected primary or secondary cells may also include a DNA sequence encoding a selectable marker which confers a selectable phenotype upon them, facilitating  
10 their identification and isolation. Methods for producing transfected primary, secondary cells which stably express the DNA sequence, clonal cell strains and heterogenous cell strains of such transfected cells, methods of producing the clonal and heterogenous cell strains, and methods of treating or preventing an abnormal or undesirable condition through the use of populations of transfected primary or secondary cells are part of the invention.

15

#### Transfection of Primary or Secondary Cells, Homologous Recombination and Production of Clonal or Heterogenous Cell Strains

Vertebrate tissue can be obtained by standard methods such as punch biopsy or other surgical methods of obtaining a tissue source of the primary cell type of interest. For  
20 example, punch biopsy is used to obtain skin as a source of fibroblasts or keratinocytes. A mixture of primary cells is obtained from the tissue, using known methods, such as enzymatic digestion or explanting. If enzymatic digestion is used, enzymes such as collagenase, hyaluronidase, dispase, pronase, trypsin, elastase and chymotrypsin can be used.

The resulting primary cell mixture can be transfected directly or it can be cultured  
25 first, removed from the culture plate and resuspended before transfection is carried out. Primary cells or secondary cells are combined with the DNA sequence to be introduced into their genomes which optionally includes DNA encoding a selectable marker, and treated in order to accomplish transfection. In addition, the primary or secondary cells are combined with a Rad52 protein or fragment thereof and a Ku-inactivating molecule, e.g., an anti-Ku  
30 antibody, either alone or as part of a complex.

Transfected primary or secondary cells, can be made by electrophoration. Electrophoration is carried out at appropriate voltage and capacitance (and corresponding time constant) to result in entry of the DNA construct(s) into the primary or secondary cells. Electroporation can be carried out over a wide range of voltages (e.g., 50 to 2000 volts) and  
35 corresponding capacitance. Total DNA of approximately 0.1 to 500  $\mu$ g is generally used.

Preferably, primary or secondary cells are transfected using microinjection. Alternatively, known methods such as calcium phosphate precipitation, modified calcium

5 phosphate precipitation and polybrene precipitation, liposome fusion and receptor-mediated gene delivery can be used to transfect cells. A stably, transfected cell is isolated and cultured and subcultivated, under culturing conditions and for sufficient time, to propagate the stably transfected secondary cells and produce a clonal cell strain of transfected secondary cells. Alternatively, more than one transfected cell is cultured and subcultured, resulting in  
10 production of a heterogenous cell strain.

After transfection, the cell is maintained under conditions which permit homologous recombination, as is known in the art (Capecchi (1989) *Science* 244:1288-1292).

Homologously recombinant primary or secondary cells can undergo a sufficient number of doublings to produce either a clonal cell strain or a heterogenous cell strain of  
15 sufficient size to provide the therapeutic protein to an individual in effective amounts. In general, for example, 0.1 cm<sup>2</sup> of skin is biopsied and assumed to contain 100,000 cells; one cell is used to produce a clonal cell strain and undergoes approximately 27 doublings to produce 100 million homologously recombinant secondary cells. If a heterogenous cell strain is to be produced from an original homologously recombinant population of approximately  
20 100,000 cells, only 10 doublings are needed to produce 100 million cells.

The number of required cells in a homologously recombinant clonal or heterogenous cell strain is variable and depends on a variety of factors, including but not limited to, the use of the homologously recombinant cells, the functional level of the exogenous DNA sequence in the cells, the functional level of altered DNA sequence in the cell, the site of implantation  
25 of the homologously recombinant cells (for example, the number of cells that can be used is limited by the anatomical site of implantation), and the age, surface area, and clinical condition of the patient. To put these factors in perspective, to deliver therapeutic levels of human growth hormone in an otherwise healthy 10 kg patient with isolated growth hormone deficiency, approximately one to five hundred million homologously recombinant fibroblasts  
30 would be necessary (the volume of these cells is about that of the very tip of the patient's thumb).

Several methods can be used to determine the efficacy of the methods described herein to enhance homologous recombination in a cell. For example, an experimental system can be designed to detect a non-conservative substitution in a cell, e.g., a human cell. The  
35 substitution can be a C to T substitution at the CGA codon of exon 3 of the HPRT gene, which is part of an *XhoI* site. This mutation creates a TGA termination signal which results in a HPRT-negative phenotype scored as resistant to 6-thioguanine (6-TG). This mutation is

5 also accompanied by a loss of the corresponding *XhoI* site. Briefly, a DNA sequence which includes the C to T substitution can be introduced by microinjection into a human fibroblast cell as part of a complex which includes an agent which enhances homologous recombination and an agent which inactivates Ku. The cells are cultured and allowed to propagate prior to introducing the cells onto media which includes 6-TG. 6-TG resistant clones can then be  
10 scored to determine the presence of the mutated DNA sequence. The presence of a homologous recombination event can be detected by Southern blot analysis of *XhoI* digested genomic DNA using an HPRT-specific probe. The results can also be compared to control cells in which the mutated DNA sequence is introduced in the absence of an agent which enhances homologous recombination and an agent which inactivates Ku.

15  
Implantation of Clonal Cell Strain or Heterogenous Cell Strain of Homologously Recombinant Secondary Cells

20 The homologously recombinant cells produced as described above can be introduced into an individual to whom the therapeutic protein is to be delivered, using known methods. The clonal cell strain or heterogenous cell strain is then introduced into an individual, using known methods, using various routes of administration and at various sites (e.g., renal subcapsular, subcutaneous, central nervous system (including intrathecal), intravascular,  
25 intrahepatic, intrasplanchnic, intraperitoneal (including intraomental), or intramuscular implantation). Once implanted in the individual, the homologously recombinant cells produce the therapeutic product encoded by the exogenous synthetic DNA or the homologously recombinant cells express a therapeutic protein encoded by an endogenous DNA sequence under the control of an exogenous regulatory sequence. For example, an  
30 individual who has been diagnosed with Hemophilia B, a bleeding disorder that is caused by a deficiency in Factor IX, a protein normally found in the blood, is a candidate for a gene therapy cure. The patient has a small skin biopsy performed; this is a simple procedure which can be performed on an out-patient basis. The piece of skin, approximately the size of a matchhead, is taken, for example, from under the arm and requires about one minute to  
35 remove. The sample is processed, resulting in isolation of the patient's cells (in this case, fibroblasts) and genetically engineered to produce the missing Factor IX. Based on the age, weight, and clinical condition of the patient, the required number of cells are grown in large-



5 scale culture. The entire process should require 4-6 weeks and, at the end of that time, the appropriate number of genetically-engineered cells are introduced into the individual, once again as an outpatient (e.g., by injecting them back under the patient's skin). The patient is now capable of producing his or her own Factor IX and is no longer a hemophiliac.

10 A similar approach can be used to treat other conditions or diseases. For example, short stature can be treated by administering human growth hormone to an individual by implanting primary or secondary cells which express human growth hormone.

As this example suggests, the cells used will generally be patient-specific genetically-engineered cells. It is possible, however, to obtain cells from another individual of the same species or from a different species. Use of such cells might require administration of an  
15 immunosuppressant, alteration of histocompatibility antigens, or use of a barrier device to prevent rejection of the implanted cells.

For many diseases, this will be a one-time treatment and, for others, multiple gene therapy treatments will be required.

Transfected primary or secondary cells can be administered alone or in conjunction  
20 with a barrier or agent for inhibiting immune response against the cell in a recipient subject. For example, an immunosuppressive agent can be administered to a subject inhibit or interfere with normal response in the subject. Preferably, the immunosuppressive agent is an immunosuppressive drug which inhibits T cell/or B cell activity in a subject. Examples of such immunosuppressive drugs commercially available (e.g., cyclosporin A is commercially  
25 available from Sandoz Corp. East Hanover, NJ).

An immunosuppressive agent e.g., drug, can be administered to a subject at a dosage sufficient to achieve the desired therapeutic effect (e.g., inhibition of rejection of the cells). Dosage ranges for immunosuppressive drugs are known in the art. *See, e.g.,* Freed et al. (1992) *N. Engl. J. Med.* 327:1549; Spencer et al. (1992) *N. Engl. J. Med.* 327:1541; Widner  
30 et al. (1992) *n. Engl. J. Med.* 327:1556). Dosage values may vary according to factors such as the disease state, age, sex, and weight of the individual.

Another agent which can be used to inhibit T cell activity in a subject is an antibody, or fragment of derivative thereof. Antibodies capable of depleting or sequestering T cells *in vivo* are known in the art. Polyclonal antisera can be used, for example, anti-lymphocyte  
35 serum. Alternatively, one or more monoclonal antibodies can be used. Preferred T cell depleting antibodies include monoclonal antibodies which bind to CD2, CD3, CD4, CD8, CD40, CD40, ligand on the cell surface. Such antibodies are known in the art and are

5 commercially available, for example, from American Type Culture Collection. A preferred antibody for binding CD3 on human T cells is OKT3 (ATCC CRL 8001).

An antibody which depletes, sequesters or inhibits T cells within a recipient subject can be administered in a dose for an appropriate time to inhibit rejection of cells upon transplantation. Antibodies are preferably administered intravenously in a pharmaceutically acceptable carrier of diluent (e.g., saline solution).

Another way of interfering with or inhibiting an immune response to the cells in a recipient subject is to use an immunobarrier. An "immunobarrier" as used herein, refers to a device which serves as a barrier between the administered cell and cells involved in immune response in a subject. For example, the cells can be administered in an implantable device.

15 An implantable device can include the cells contained within a semi-permeable barrier, i.e., one which lets nutrients and the product diffuse in and out of the barrier but which prevents entry or larger immune system components, e.g., antibodies or complement. An implantable device typically includes a matrix, e.g., a hydrogel, or core in which cells are disposed.

20 Optionally, a semi permeable coating can enclose the gel. If disposed within the gel core, the administered cells should be sequestered from the cells of the immune system and should be cloaked from the cells and cytotoxic antibodies of the host. Preferably, a permselective coating such as PLL or PLO is used. The coating often has a porosity which prevents components of the recipient's immune system from entering and destroying the cells within the implantable device.

25 Many methods for encapsulating cells are known in the art. For example, encapsulation using a water soluble gum to obtain a semi-permeable water insoluble gel to encapsulate cells for production and other methods of encapsulation are disclosed in U.S. patent No: 4,352,883. Other implantable devices which can be used are disclosed in U.S. Patent No.: 5,084,350, U.S. Patent No. 5,427,935, WO 95/19743 published July 27, 1995, 30 U.S. Patent No.: 5,545,423, U.S. Patent Number 4,409,331, U.S. Patent Number 4,663,286, and European Patent No. 301,777.

#### Uses of Homologously Recombinant Primary and Secondary Cells and Cell Strains

35 Homologously recombinant primary or secondary cells or cell strains have wide applicability as a vehicle or delivery system for therapeutic proteins, such as enzymes, hormones, cytokines, antigens, antibodies, clotting factors, anti-sense RNA, regulatory proteins, transcription proteins, receptors, structural proteins, novel (non-optimized) proteins

5 and nucleic acid products, and engineered DNA. For example, homologously recombinant primary or secondary cells can be used to supply a therapeutic protein, including, but not limited to, erythropoietin, calcitonin, growth hormone, insulin, insulinotropin, insulin-like growth factors, parathyroid hormone,  $\alpha$ 2-interferon (IFNA2),  $\beta$ -interferon,  $\gamma$ -interferon, nerve growth factors, FSH $\beta$ , TGF- $\beta$ , tumor necrosis factor, glucagon, bone growth factor-2, bone growth factor-7, TSH- $\beta$ , interleukin 1, interleukin 2, interleukin 3, interleukin 6, interleukin 10 11, interleukin 12, CSF-granulocyte (GCSF), CSF-macrophage, CSF-granulocyte/macrophage, immunoglobulins, catalytic antibodies, protein kinase C, glucocerebrosidase, superoxide dismutase, tissue plasminogen activator, urokinase, antithrombin III, DNase,  $\alpha$ -galactosidase, tyrosine hydroxylase, blood clotting factor V, 15 blood clotting factor VII, blood clotting factor VIII, blood clotting factor IX, blood clotting factor X, blood clotting factor XIII, apolipoprotein E, apolipoprotein A-I, globins, low density lipoprotein receptor, IL-2 receptor, IL-2 antagonists,  $\alpha$ -1-antitrypsin, immune response modifiers,  $\beta$ -glucoceramidase,  $\alpha$ -iduronidase,  $\alpha$ -L-iduronidase, glucosamine-N-sulfatase,  $\alpha$ -N-acetylglucosaminidase, acetylcoenzymeA: $\alpha$ -glucosamine-N-acetyltransferase, 20 N-acetylglucosamine-6-sulfatase,  $\beta$ -galactosidase,  $\beta$ -glucuronidase, N-acetylgalactosamine-6-sulfatase, and soluble CD4.

All patents and references cited herein are incorporated in their entirety by reference. Other embodiments are within the following claims.

25

5

What is claimed:

1. A complex for promoting alteration of a target sequence in a cell, comprising:  
a double stranded DNA sequence, a homologous recombination-enhancing agent, and an  
10 agent which inhibits non-homologous end joining.
2. The complex of claim 1, wherein the homologous recombination-enhancing  
agent is selected from the group consisting of: a Rad52 protein or functional fragment  
thereof, a Rad51 protein or functional fragment thereof, a Rad54 protein or functional  
15 fragment thereof, and combinations thereof.
3. The complex of claim 1, wherein the homologous recombination-enhancing  
agent is Rad52 protein or functional fragment thereof.
- 20 4. The complex of claim 1, wherein the agent which inhibits non-homologous  
end joining is selected from the group consisting of an agent which inactivates hMre11, an  
agent which inactivates hRad50, an agent which inactivates Nbs1, an agent which inactivates  
hLig4, an agent which inactivates hXrcc4, and an agent which inactivates Ku.
- 25 5. The complex of claim 1, wherein the agent which inhibits non-homologous  
end joining is a Ku inactivating agent.
6. The complex of claim 5, wherein the Ku-inactivating agent is selected from  
the group consisting of: an anti-Ku antibody, a Ku-binding oligomer, and a Ku-binding  
30 polypeptide.
7. The complex of claim 5, wherein the Ku-inactivating agent is an anti-Ku-  
antibody.
- 35 8. The complex of claim 1, wherein the DNA sequence comprises a linear DNA  
sequence.



5           9.       The complex of claim 1, wherein the DNA sequence is flanked by at least one targeting sequence.

          10.       The complex of claim 1, wherein the DNA sequence comprises an exogenous regulatory sequence.

10           11.       The complex of claim 10, wherein the regulatory sequence is a promoter, an enhancer, an upstream activating sequence, a scaffold-attachment region or a transcription factor-binding site.

15           12.       The complex of claim 11, wherein the regulatory sequence is a promoter and an enhancer.

          13.       The complex of claim 11, wherein the regulatory sequence is a promoter and an upstream activating sequence.

20           14.       The complex of claim 3, wherein the Rad52 protein or functional fragment thereof is coated on the DNA sequence.

          15.       The complex of claim 3, wherein the Rad52 protein or fragment thereof is  
25 human Rad52.

          16.       The complex of claim 7, wherein the anti-Ku antibody is an anti-Ku70 antibody.

30           17.       The complex of claim 7, wherein the anti-Ku antibody is an anti-Ku80 antibody.

          18.       The complex of claim 7, wherein at least one anti-Ku antibody is covalently linked to the DNA sequence.

35           19.       The complex of claim 7, wherein at least one anti-Ku antibody is covalently linked to the agent for enhancing homologous recombination.

5

20. The complex of claim 7, wherein the complex comprises an anti-Ku70 antibody and an anti-Ku80 antibody.

10

21. The complex of claim 9, wherein the DNA sequence comprises a regulatory sequence and the targeting sequence is derived from a sequence 5' of the protein coding region of the FSH $\beta$  gene.

15

22. The complex of claim 9, wherein the DNA sequence comprises a regulatory sequence and the targeting sequence is derived from a sequence 5' of the protein coding region of the IFN $\alpha$  gene.

20

23. The complex of claim 9, wherein the DNA sequence comprises a regulatory sequence and the targeting sequence is derived from a sequence 5' of the protein coding region of the GCSF gene.

25

24. The complex of claim 1, further comprising an agent which inhibits a mismatch repair protein.

25. A method of promoting an alteration at a selected site in a target DNA of a cell, comprising:

introducing into the cell a double stranded DNA sequence, an agent which enhances homologous recombination, and an agent which inhibits non-homologous end joining, to thereby promote alteration of the chromosomal DNA, to thereby promote alteration at a selected site in the chromosomal DNA.

30

26. The method of claim 25, wherein the DNA sequence comprises a linear DNA sequence.

35

27. The method of claim 25, wherein the DNA sequence is flanked by at least one targeting sequence.

5           28.     The method of claim 25, wherein the DNA sequence comprises an exogenous regulatory sequence.

          29.     The method of claim 28, wherein the regulatory sequence selected from the group consisting of: a promoter, an enhancer, an upstream activating sequence, a scaffold-  
10 attachment region and a transcription factor-binding site.

          30.     The method of claim 28, wherein the regulatory sequence is a promoter and an enhancer.

15           31.     The method of claim 28, wherein the regulatory sequence is a promoter and an upstream activating sequence.

          32.     The method of claim 25, wherein the agent which enhances homologous recombination is selected from the group consisting of: a Rad52 protein or functional  
20 fragment thereof, a Rad51 protein or functional fragment thereof, a Rad54 protein or functional fragment thereof, and combinations thereof.

          33.     The method of claim 25, wherein the agent which enhances homologous recombination is a Rad52 protein or functional fragment thereof.

25           34.     The method of claim 33, wherein the Rad52 protein or functional fragment thereof is coated on the DNA sequence.

          35.     The method of claim 33, wherein the Rad52 protein or fragment thereof is  
30 human Rad52.

          36.     The method of claim 25, wherein the agent which inhibits non-homologous end joining is selected from the group consisting of an agent which inactivates hMre11, an agent which inactivates hRad50, an agent which inactivates Nbs1, an agent which inactivates  
35 hLig4, an agent which inactivates hXrcc4, and an agent which inactivates Ku.

5           37.     The method of claim 25, wherein the agent which inhibits non-homologous end joining is a Ku inactivating agent.

          38.     The method of claim 37, wherein the agent which inactivates Ku is an anti-Ku antibody, a Ku-binding oligomer, and a Ku-binding polypeptide.

10           39.     The method of claim 37, wherein the agent which inactivates Ku is a Ku antisense molecule.

          40.     The method of claim 37, wherein the agent which inactivates Ku is an anti-Ku  
15     antibody.

          41.     The method of claim 40, wherein the anti-Ku antibody is an anti-Ku70 antibody.

20           42.     The method of claim 40, wherein the anti-Ku antibody is an anti-Ku80 antibody.

          43.     The method of claim 40, wherein at least one anti-Ku antibody is covalently linked to the DNA sequence.

25           44.     The method of claim 40, wherein at least one anti-Ku antibody is covalently linked to the Rad52 protein or fragment thereof.

          45.     The method of claim 25, wherein the cell is of fungal, plant or animal origin.

30           46.     The method of claim 45, wherein the cell is of vertebrate origin.

          47.     The method of claim 46, wherein the cell is a primary or secondary mammalian cell.

35           48.     The method of claim 46, wherein the cell is a primary or secondary human cell.



5

49. The method of claim 46, wherein the cell is an immortalized mammalian cell.

50. The method of claim 46, wherein the cell is an immortalized human cell.

10

51. The method of claim 25, wherein the DNA sequence, the agent which enhances homologous recombination and the agent which inhibits non-homologous end joining are introduced into the cell as a complex.

15

52. The method of claim 25, further comprising introducing an agent which inhibits a mismatch-repair protein.

53. The method of claim 52, wherein the mismatch-repair protein is selected from the group consisting of: Msh2, Msh6, Msh3, Mlh1 and PMS2.

20

54. The method of claim 52, wherein the agent is an agent which inhibits expression of a mismatch-repair protein.

55. The method of claim 54, wherein the agent is an anti-mismatch-repair protein antibody.

25

56. The method of claim 54, wherein at least one anti-mismatch-repair protein antibody is covalently linked to the DNA sequence.

30

57. The method of claim 55, wherein at least one anti-mismatch-repair protein antibody is covalently linked to the Rad52 protein or fragment thereof.

58. The method of claim 27, wherein the DNA sequence comprises a regulatory sequence and the targeting sequence is derived from the region 5' of the FSH $\beta$  coding region.

35

59. The method of claim 27, wherein the DNA sequence comprises a regulatory sequence and the targeting sequence is derived from the region 5' of the IFN $\alpha$  coding region.

5           60.     The method of claim 27, wherein the DNA sequence comprises a regulatory sequence and the targeting sequence is derived from the region 5' of the GCSF coding region.

          61.     The method of claim 25, wherein the target DNA comprises a mutation having less than 10 base pairs which differ from wild-type sequence.

10           62.     The method of claim 61, wherein the mutation is a point mutation.

          63.     The method of claim 62, wherein the DNA sequence comprises a wild-type sequence which can correct the mutation.

15           64.     The method of claim 63, wherein the target DNA is a cystic fibrosis transmembrane regulator (CFTR) gene.

          65.     The method of claim 64, wherein the mutation changes an amino acid encoded  
20 by codon 508 of the coding region of the CFTR gene.

          66.     The method of claim 63, wherein the target DNA is a  $\beta$ -globin gene.

          67.     The method of claim 66, wherein the mutation changes an amino acid encoded  
25 by codon 6 of the coding region of the  $\beta$ -globin gene.

          68.     The method of claim 63, wherein the target DNA is a Factor VIII gene.

          69.     The method of claim 68, wherein the mutation changes an amino acid encoded  
30 by codon 2209 of the coding region of the Factor VIII gene.

          70.     The method of claim 68, wherein the mutation changes an amino acid encoded by codon 2229 of the coding region of the Factor VIII gene.

35           71.     The method of claim 63, wherein the target DNA is a Factor IX gene.

5           72.     The method of claim 63, wherein the target DNA is a von Willebrand factor gene.

          73.     The method of claim 63, wherein the target DNA is a xeroderma pigmentosa group G gene.

10           74.     A homologously recombinant cell made by the method of claim 25.

          75.     A method of altering expression of a protein coding sequence of a gene in a cell, the method comprising:

15           introducing the complex of claim 1 into the cell, wherein the DNA sequence comprises a regulatory sequence;

                  maintaining the cell under conditions which permit alteration of a targeted genomic sequence to produce a homologously recombinant cell; and

                  maintaining the homologously recombinant cell under conditions which  
20           permit expression of the protein coding sequence of the gene under control of the regulatory sequence, thereby altering expression of the protein coding sequence of the gene.

1/11

## SEQUENCE LISTING

&lt;110&gt; Evguenii Ivanov

&lt;120&gt; METHODS OF IMPROVING HOMOLOGOUS RECOMBINATION

&lt;130&gt; 10278/016001

&lt;160&gt; 9

&lt;170&gt; FastSEQ for Windows Version 3.0

&lt;210&gt; 1

&lt;211&gt; 7622

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 1

ggatccgaga	acatagaagg	agcaggtaat	ttatcaagge	atgaacacgg	gtgcttaatt	60
tcctattttg	aggccaggca	tgggtggctca	cacctgtaat	cccaacactt	taggaagcca	120
aggtgggtgg	attgcttgag	tctaggattt	tgagaccagc	ctggccaaca	tggcgaaatc	180
ctgtctctac	taaaaatact	aaaattaacc	agtcattggtg	gtggtgtgcc	tttagtcca	240
gctactctgg	tggctgaggg	acaagaatca	cttgaacctg	ggaggcagag	gttgcagtga	300
gctgagactg	tgccacttca	ctccagcctg	ggtgacagag	taagattctg	tctcaaaaaa	360
tatgtatata	tacacacata	taatagatac	ataaacatat	atacatatat	aatatatata	420
tatatatatt	atatataata	tataaacata	tataaatata	tatatatata	tatatatata	480
tatataaacc	aaacataaag	gaataatttt	gggggaaaat	cttcataaat	gaaagaacaa	540
cataggctgt	tgagtatatg	cacagaaatt	caagagatct	tccagcaatt	gaagacattg	600
gtttaccaga	attcacaaaa	gaagtcagct	gtgcatttaa	agtagaatgt	gatgagtgtt	660
accactgagg	taggaactgg	gaactaagga	agcgttaagac	agaaagtgtc	gaactgagag	720
ttgggcattg	gaggctgtgt	aaggcagggt	aagtgaatgt	ctcctagaag	ctacctttta	780
atggagtttt	gaagtacttg	taggagttagc	ttagggtgaa	agaagaggag	aaacatgtat	840
caggcagagg	gactagaacc	ttattacctt	caaagaagaa	gcaaaaagaa	tacatgtgac	900
tttgagggtg	tgggaggtgc	tttaagccaa	tatagggtgaa	tttgacatag	gacttcccta	960
aataatgttc	ggtcattttgt	taaatattga	gtgatataatc	actgtattaa	agcccaagag	1020
ttgctttttat	atagaaagaa	gaaaaaagcc	caagagagtt	ttattttctag	agggaaatatt	1080
ttctagaaat	aaaggaaggt	gtatcagcca	gtttctagtc	aggaaaacag	aatcacacc	1140
tgatatgcaa	aatagaggaa	aatcagggaa	ttcattaatc	cagagatttg	gttgctcaag	1200
tattagattg	ctgaaaagcc	agacagggaa	tatgaggcaa	tcagagataa	gtattagtga	1260
caagctccat	ttatgtgcag	gattggaggg	acatagggtg	ggttcccaga	agccagaagg	1320
tgagaccacc	tagcagaagc	tcaaaccaca	gctgggggtt	cctcacaaaa	gctgggacca	1380
ccaggaggag	ctgtccaatg	ggatctggag	ccaggggagat	catgcagtc	ctaccaggaa	1440
gggaagcaga	atgtaaaagg	tagagagaaa	tactccaact	gcttccttgc	attcactttc	1500
caatctccat	tcacaaaggc	aaaaacctgc	taatacagca	gagtgggaaa	agcagcctgc	1560
caaggtcctt	tctcccacaa	aacagagcac	aaaaccaagc	aaaaacaagg	aatgcatttg	1620
atagcaaaca	ggctatggac	caacccaaca	taaaagaaat	gatgagtgat	ttcttttttc	1680
atgttggttca	agaaaagtat	ttcagtaact	attatgtaac	agaaattcta	tttatttttg	1740
ggaattcaaa	ggtgaataaa	aaagaactct	aaatttttat	caataaaata	tttcaaaaac	1800
ctcaatgaga	gtaatggcat	taactagcaa	atatgcta	gagatgagct	agccataaga	1860
ggcttagaat	tgagagaaag	gtctgggggc	ctcttgacag	gccaaattca	gagctgtttg	1920
tgggaatctc	tgacctaa	gcagggtggaa	atataaatat	gggcatttag	aatagtggcc	1980
caaacttttg	atgattttctg	tcttggggtc	tctccaatta	atgggattga	tgagaactgt	2040
agaccactga	ggtcaccatg	gctcaatgaa	tagtcccctg	gctttggagt	caaactgacc	2100
tgaatatgaa	ccccagcttt	gctacttaca	ggttgcat	atcctcagtt	ttctcatctt	2160
tcaaagaaga	acagtaactt	ctttaaaagg	ttattgtagg	ctgggtgcag	tggctcacgc	2220
ctgtaatcgc	agcacttttg	gaggcggagg	ctagtggatc	acttgaggcc	aggagtggga	2280
aactagcctg	gccaacatgg	tgaaactctg	tctctacaaa	aagaaattta	aaaaattttg	2340
ctgggtgtgg	tggcacacac	ctggaattcc	agctacctgg	gaggccgagg	catgagcatc	2400
acttgagtct	ggaaagcaga	gggttgca	gagccaagat	tgtaccactg	tactcaagcc	2460
tgggtgacac	agtgagacct	tgtctaaaaa	aaaaaaagg	tattgtgtta	ttgtaaatat	2520
tgtatatgaa	cttctattta	acatgttttag	ttaaatgcct	gtgtaattgt	ccaatgtgct	2580



2/11

cttctagctc	actgcacaga	caaaactgat	tcactgaaat	catggaattg	cagcaaagaa	2640
caaatactaat	taatgtaggt	caaacgggag	gactggagtt	attattcaaa	tcagtctccc	2700
tgaaaactca	gaggctaggg	ttttatggat	aatttggttg	gcaggggact	aggggaatggg	2760
tgctgctgat	tgggtgggga	atgaaatagt	aagattgttg	aaaactgtcc	tccttcattg	2820
agtctgcttc	cgggtgtagg	ccacacgacc	agttgagtca	tgaagcatgc	gtccaagtgg	2880
agtcagtttg	ttgccagaat	gcaaaagcct	gaaaaatgtc	tcaaatgac	aactgtaggc	2940
tccacaataa	tgatattatc	tataggagca	attggggaag	taacaaatct	tgtgacctct	3000
ggacacataa	ctcctgaact	agtaagggat	tataaaaacc	atgcctatat	cttatcagaa	3060
ttcaggtccc	cccataatcc	taatctcaca	gcatttcatt	tgtttagaaa	ggccatttttc	3120
agtccttgag	caaggagggg	gttagtttta	ggataggact	attatccttg	cttcgttaaa	3180
ctataaacta	aattcctccc	atggttagct	tggcctacac	ctaagaatga	gtgagaacag	3240
ccagcctgtg	aggctagagg	caagatggag	tcagccatgc	tagatttatc	tcactgtcat	3300
aacctttgca	aaggcagttt	cacctgggac	ataggaggta	ctcaatgaaa	aagaagctat	3360
taatattaaa	atttttaaaa	tgaatttaag	gaactaatac	tatgtacata	ttagtcatta	3420
aaacaaagtg	gttcattttac	attcacacaa	ataaatcttg	tgattataca	taggtaatat	3480
gaaaaacttt	gttttctttc	ataatacaag	gtattagcaa	tagatatagt	aatgttagca	3540
ttcctttgga	aaaaatgaaa	agattttataa	ttttccaaga	atcattagta	tttttattta	3600
atatacataa	tataaaattt	attcattcta	taacttggaa	atatgcttgc	ttaccaatta	3660
ctgacagatt	tcaaaatatt	tctatactca	caatattcat	ttacataaat	attgatttgg	3720
tacttacaaat	gtgtactgct	atgctaagtt	ttgtctttgt	caaacatatt	ttataaaatc	3780
ataatcctag	atgaatccaa	ctttttggtaa	cccacgtgcc	tgaacccctg	ctgttaacag	3840
gcaaagtgtg	gtaggtacag	atctatacct	accaccttcc	tctacccacc	agcatctgca	3900
cccaccaccc	ctccccaccc	accattatct	ataccaacca	cccctcccaa	cctaccagca	3960
tctgcaccca	ccacacogcc	cacccaccac	catgtacact	cactacacct	tccagccatc	4020
accatctgca	cccatcactc	ctccccatcc	acaagcatct	gcacccacca	catttcccta	4080
cctaccagca	tcttcactca	ccacctctcc	accaccagc	atctgcaccc	acaacccctc	4140
ctcaccaccc	agagtctgca	tccatcacac	ttgcccactc	gctagcatct	gcaccatcaa	4200
gctctgcctt	cttgcctaata	acgggatgag	ctctccatgg	ttctgcctaa	agacaatgct	4260
tccactcctc	ttctataacc	catttccttt	tacctcttca	agtacacttc	agaacttctc	4320
tctccttctg	ataccaactt	tttccacttt	actcaatcat	tctatcacc	atacaaactg	4380
gtttattttct	cccattcttaa	agttaaaaat	caaaagaaaa	ttgtctgcgg	ccaggcacgg	4440
tggtcacgc	ctgtaatccc	aacacttttg	gaggccaagg	agggttggat	gacttaaggt	4500
taggagttca	agaccagcct	ggccaacatg	gtgaaaccca	tctctactaa	aaatacaaaa	4560
attagccagg	catggtggca	catgcctgta	gtctcaggta	cttgggaggc	tgaggccaga	4620
gaatggcttg	aaccogggag	gcagaggttg	cagtgcagccg	agattgtgcc	cttgcactcc	4680
agcctgggtg	acagagtggg	actccatctc	aaaaataaaa	aataaaaaata	aaacaaaaga	4740
aagttatttt	taccaaacat	ccacattaac	caaataccca	tttctttatt	gatctttgta	4800
aaaaaaagct	cttggaaaaa	ttgtctatat	tcactatgac	ttatctcctc	caaatcactt	4860
aaacacatac	caatcaggtt	tttgttttca	tcattccaaa	gtaactttta	cagccaagga	4920
cagtagcgaa	ctttacatcg	catatgcatt	gtgaagttct	tgatcctcat	cttacttaac	4980
ctgtcagcag	tatctgacac	aggtgtcact	ggctcctccc	tgagatgctc	tctttatttg	5040
gctttgggga	caccatattc	tccccattcc	tactttcctc	aatggccctc	ctcagtctcc	5100
tttggaagaa	ggaaaaagaa	acttcattat	ctcctggatg	tagtacaac	aactcaagct	5160
caacatgtgc	atactgaact	ccatttcctt	ttcccaaact	tcgacattta	cagccatccc	5220
ctttcagctg	atagcaagtt	tatccttcca	gctactcaaa	ccagaatctt	tagagccatc	5280
cttgaccctt	ttcctcctct	cacactcaac	atctatccat	cagaaaatct	tgttggttct	5340
actttcaaaa	tgcatacaga	gtcagagcat	gtctcattac	ctccaatagc	taccatacta	5400
gtctgaacaa	acatcatttc	tcacctgggt	tattgaacaa	acatcatttc	tcacctgggt	5460
tattgatagc	atcctaacgg	gtcttcctgt	ttcttggttc	ccctatatata	gcaacacagc	5520
agtcagagga	gtccttttag	aactcaatca	gatcatgtca	cgctactcct	ctacttaaaa	5580
tccttcaatg	ggtcccatta	cacaaagagt	acaaaccaga	gcccttacac	tggtctacaa	5640
gttccaacat	ttgactcctg	ttatctctct	gacatcatat	tctaataatta	ctgctgttgt	5700
ccttttgctc	cagtcacact	gttttgattag	taaatattta	ttaaacaaag	caatcctagt	5760
ctccaaagag	atcatagttt	attggaggaa	acaagagcct	ataaatgggt	acacacagaa	5820
ggtagtgatt	atgggttctc	ctcacctccc	atcctaaact	ttgacagggt	aaactcccct	5880
ggatgttgaa	ggttgaggaa	tttgccaggg	ttcagggttg	tgttgaggga	ggcaggggag	5940
aagcaaggac	atttcaggga	ggaagaacat	tacatgcaaa	gatctaaaga	tatgaatcag	6000
caacatattt	atggaattac	aagtaaagta	gaaagtctct	gctaaaacat	caaaaaataa	6060
agatttgatga	ttagggggcc	agaatgtggg	agggaaagag	agatacagtt	cacactttta	6120
gacaggagcc	agatcatgaa	atgttttctc	tttgtttgtt	tcttccctca	cagcttttga	6180
tatgctcttg	gagcaattta	ttaaccatat	tttttaatgc	atctcctgaa	cagagtcaaa	6240
gcaatacttg	gaaaggactc	tgaatttctc	gatttaaaga	tacaaaagaa	aaatctggag	6300

3/11

tcacaatttaa	tttgagaagg	taaaggagtg	ggtgtgctac	tgtatcaaata	ttaatttgta	6360
caaaatcatc	atctctagta	acattatttt	ttctaatacta	ctgcgttttag	actacttttag	6420
taaagcctga	tctccctgtc	tatctaaaca	ctgattcact	tacagcaagc	ttcaggctag	6480
cattgggtcat	attaataccc	aacaaatcca	caaggtgtta	gttgcacatg	attttgtata	6540
aaaggtgaac	tgagatttca	ttcagtctac	agctcttgcc	aggcaaggca	gccgaccaca	6600
ggtgagtcct	ggcatctacc	gttttcaagt	gtgacagcta	cttttgaaat	tacagatttg	6660
tcaggacatg	gaggacaaaa	ctagagcttc	tcactactgt	tgtgtaggaa	atttatgctt	6720
gtcaacctgg	cttgtaaaaat	atggttaata	taacgtaatc	actgttagca	agtaactgac	6780
tttatagacc	aatatgcctc	tcttctgaaa	tggtcttatt	ttaaacaaat	gtgagcaaaa	6840
gaaaatattt	atgagattct	aaaaatgaag	acataatttt	gtagtataga	attttcttgg	6900
ccaggaatgg	tggctcatgc	ttgtaatccc	agcacttttg	gaggccaagg	tcagaggatt	6960
gcttgagcct	ggaagggtga	agatgcagtg	attcatgatt	ataccactgc	actccagcct	7020
gggcaacaga	gcaagaccct	gtctcaagaa	aagaaaagaa	ttttattttt	cttttcagac	7080
aaaaatagac	tttaaaataa	taatggaaga	acaaatatga	tgatcacaat	tatcagagta	7140
attactttat	gacagtcagc	aataagattc	taatctttaa	atattcctct	gcttaaatca	7200
ttatatggga	gttttgatct	ataatatatt	cccaccctga	cccaaaaatt	gaagaaggac	7260
aaggaaaaat	gttggttcaa	gaaacaaaga	tgtaagtaaa	aaggcataag	gaaggaaaaa	7320
aaacttttga	agcaaaatgt	gattgaggag	gatgagcaga	ccaattattt	ttggtttggt	7380
cagcttacat	aatgattatc	gttctttggt	ttctcagttt	ctagtgggct	tcattgtttg	7440
cttcccagac	caggatgaag	acactccagt	ttttcttcct	tttctgttgc	tggaaagcaa	7500
tctgctgcaa	tagctgtgag	ctgaccaaca	tcaccattgc	aatagagaaa	gaagaatgtc	7560
gtttctgcat	aagcatcaac	accacttggt	gtgctggcta	ctgctacacc	agggtaggta	7620
cc						7622

&lt;210&gt; 2

&lt;211&gt; 6038

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 2

ggatccgaga	acatagaagg	agcaggtaat	ttatcaaggc	atgaacacgg	gtgcttaatt	60
tcctattttg	aggccaggca	tggtggctca	cacctgtaat	cccaacactt	taggaagcca	120
aggtgggtgg	attgcttgag	tctaggattt	tgagaccago	ctggccaaca	tggcgaaatc	180
ctgtctctac	taaaaatact	aaaattaacc	agtcattggtg	gtggtgtgcc	tttagtccca	240
gctactctgg	tggctgaggc	acaagaatca	cttgaacctg	ggaggcagag	gttgcagtga	300
gctgagactg	tgccacttca	ctccagcctg	ggtgacagag	taagattctg	tctcaaaaaa	360
tatgtatata	tacacacata	taatagatac	ataaacatat	atacatatat	aatatataaa	420
tatatatatt	atatataata	tataaacata	tataaatata	tatatatata	tatatatata	480
tatataaacc	aaacataaag	gaataatttt	gggggaaaat	cttcataaat	gaaagaacaa	540
cataggctgt	tgagtatatg	cacagaaatt	caagagatct	tccagcaatt	gaagacattg	600
gtttaccaga	attcacaaaa	gaagtcagct	gtgcatttaa	agtagaatgt	gatgagtgtt	660
accactgagg	taggaactgg	gaactaagga	agcgtaagac	agaaagtgct	gaactgagag	720
ttgggcattg	gaggctgtgt	aaggcagggg	aagtgaatgt	ctcctagaag	ctacctttaa	780
atggagtttt	gaagtacttg	taggagtagc	ttagggtgaa	agaagaggag	aaacatgtat	840
caggcagagg	gactagaacc	ttattacctt	caaagaagaa	gcaaaaagaa	tacatgtgac	900
tttgagggtg	tgggagggtgc	tttaagccaa	tatagggtgaa	tttgacatag	gacttcccta	960
aataatgttc	ggtcattttgt	taaatatattga	gtgatataatc	actgtattaa	agcccaagag	1020
ttctagaaat	aaaggaaggt	gtatcagcca	gtttctagtc	aggaaaacag	aatcacacc	1080
tgatatgcaa	aatagaggaa	aatcagggaa	ttcattaatc	cagagatttg	gttgctcaag	1140
tattagattg	ctgaaaagcc	agacagggaa	tatgaggcaa	tcagagataa	gtattagtga	1200
caagctccat	ttatgtgcag	gattggaggg	acatagggtg	ggttcccaga	agccagaagg	1260
tgagaccacc	tagcagaagc	tcaaaccaca	gctgggggtt	cctcacaana	gctgggacca	1320
ccaggaggag	ctgtccaatg	ggatctggag	ccagggagat	catgcagtca	ctaccaggaa	1380
gggaagcaga	atgtaaaagg	tagagagaaa	tactccaact	gcttccttgc	attcactttc	1440
caatctccat	tcacaaaggc	aaaaacctgc	taatacagca	gagtgggaaa	agcagcctgc	1500
caaggtcctt	tctcccacaa	aacagagcac	aaaaccaagc	aaaaacaagg	aatgcatttg	1560
atagcaaaca	ggctatggac	caaccaca	taaaagaaat	gatgagtgat	ttcttttttc	1620
atttggttca	agaaaagtat	ttcagtaact	attatgtaac	agaaattcta	tttattttgg	1680
ggaattcaaa	ggtgaataaa	aaagaactct	aaatttttat	caataaaaata	tttcaaaaac	1740
ctcaatgaga	gtaatggcat	taactagcaa	atatgcta	gagatgagct	agccataaga	1800
ggcttagaat	tgagagaaaag	gtctgggggc	ctcttgacag	gccaaattca	gagctgtttg	1860



4/11

tgggaatctc	tgacctaaact	gcaggtggaa	atataaatat	gggcatttag	aatagtggcc	1980
caaacttttg	atgatttctg	tcttgggggc	tctccaatta	atgggattga	tgagaactgt	2040
agaccactga	ggtcaccatg	gctcaatgaa	tagtcccctg	gctttggagt	caaactgacc	2100
tgaatatgaa	ccccagcttt	gctacttaca	ggttgcattt	atcctcagtt	ttctcatctt	2160
tcaaagaaga	acagtaactt	ctttaaaagg	ttattgtagg	ctgggtgcag	tggctcacgc	2220
ctgtaatcgc	agcacttttg	gaggcggagg	ctagtggatc	acttgaggcc	aggagttaga	2280
aactagcctg	gccaacatgg	tgaactctg	tctctacaaa	aagaaattta	aaaaattttg	2340
ctgggtgtgg	tggcacacac	ctggaattcc	agctacctgg	gaggccgagg	catgagcatc	2400
acttgagtct	ggaaagcaga	gggttgcaat	gagccaagat	tgtaccactg	tactcaagcc	2460
tgggtgacac	agtgagacct	tgtctaaaaa	aaaaaaagg	tattgtgtta	ttgtaaatat	2520
tgtatatgaa	cttctattta	acatgtttag	ttaaatgcct	gtgtaattgt	ccaatgtgct	2580
cttctagctc	actgcacaga	caaaactgat	tcactgaaat	catggaattg	cagcaaagaa	2640
caaactctaat	taatgtaggt	caaacgggag	gactggagtt	attattcaaa	tcagtctccc	2700
tgaaaactca	gaggctaggg	ttttatggat	aatttggtgg	gcaggggact	aggggaatggg	2760
tgctgctgat	tgggttgggga	atgaaatagt	aagattgtgg	aaaactgtcc	tccttcattg	2820
agtctgcttc	cgggtgtagg	ccacacgacc	agttgagtca	tgaagcatgc	gtccaagtgg	2880
agtcagtttg	ttgccagaat	gcaaaagcct	gaaaaatgtc	tcaaatgata	aactgtaggc	2940
tccacaataa	tgatattatc	tataggagca	attgggggag	taacaaatct	tgtgacctct	3000
ggacacataa	ctcctgaact	agtaagggat	tataaaaacc	atgcctatat	cttatcagaa	3060
ttcaggtccc	cccataatcc	taatctcaca	gcatttcatt	tgttttagaaa	ggccattttc	3120
agtccttgag	caaggagggg	gttaggtttt	ggataggact	attatccttg	cttcgttaaa	3180
ctataaaacta	aattcctccc	atgggttagct	tggcctacac	ctaagaatga	gtgagaacag	3240
ccagcctgtg	aggctagagg	caagatggag	tcagccatgc	tagatttata	tcactgtcat	3300
aacctttgca	aaggcagttt	cacctgggac	ataggaggta	ctcaatgaaa	aagaagctat	3360
taatatttaa	attttaaaaa	tgaatttaag	gaactaatat	tatgtacata	ttagtcatta	3420
aaacaaagtg	gttcattttac	attcacacaa	ataaatcttg	tgattataca	taggtaatat	3480
gaaaaacttt	gtttttctttc	ataatacaag	gtattagcaa	tagatatagt	aatgttagca	3540
ttccttttga	aaaaatgaaa	agatttataa	ttttccaaga	atcattagta	tttttatatta	3600
atatacataa	tataaaattt	attcattcta	taacttggaa	atatgcttgc	ttaccaatta	3660
ctgacagatt	tcaaaatatt	tctataacta	caatattcat	ttacataaat	attgatttgg	3720
tacttacaat	gtgtactgct	atgctaagtt	ttgtctttgt	caaacatatt	ttataaaatc	3780
ataatcctag	atgaatccaa	cttttggtta	cccacgtgcc	tgaacccttg	ctgttaacag	3840
gcaaagtgtg	gtaggtacag	atctatacct	accaccttcc	tctacccacc	agcatctgca	3900
cccaccaccc	ctccccaccc	accattatct	ataccaacca	cccctcccaa	cctaccagca	3960
tctgcaccca	ccacaccgcc	cacccaccac	catgtacact	cactacacct	tccagccatc	4020
accatctgca	cccatcactc	ctccccatcc	acaagcatct	gcacccacca	catttcccta	4080
cctaccagca	tcttcactca	ccacctctcc	accaccagc	atctgcaccc	acaacccctc	4140
ctcaccaccc	agagtctgca	tccatcacac	ttgcccactc	gctagcatct	gcaccatcaa	4200
gctctgcctt	cttgccataat	acgggatgag	ctctccatgg	ttctgcctaa	agacaatgct	4260
tccactcctc	ttctataacc	catttccctt	tacctcttca	agtacacttc	agaacttctc	4320
tctccttctg	ataccaactt	tttccacttt	actcaatcat	tcctatcacc	atacaaacgt	4380
gtttattttct	cccatcttaa	agttaaaaat	caaaagaaaa	ttgtctgcgg	ccaggcacgg	4440
tggctcacgc	ctgtaatccc	aacacttttg	gaggccaagg	agggttggat	gacttaaggt	4500
taggagttca	agaccagcct	ggccaacatg	gtgaaacca	tctctactaa	aaatacaaaa	4560
attagccagg	catgggtggca	catgcctgta	gtctcaggta	cttgggaggc	tgaggccaga	4620
gaatggcttg	aaccggggag	gcagagggtg	cagtgagccg	agattgtgcc	cttgcaactcc	4680
agcctgggtg	acagagttag	actccatctc	aaaaataaaa	aataaaaata	aaacaaaaga	4740
aagttatttt	tacccaacat	ccacattaac	caaataccca	tttctttatt	gatctttgta	4800
aaaaaaagct	cttggaaaaa	ttgtctatat	tcactatgac	ttatctcctc	caaatacactt	4860
aaacacatac	caatcaggtt	tttgttttca	tcattccaaa	gtaactttta	cagccaagga	4920
cagttagcgaa	ctttacatcg	catatgcatt	gtgaagttct	tgatcctcat	cttacttaac	4980
ctgtcagcag	tatctgacac	agggtgtcact	ggctcctccc	tgagatgctc	tctttatttg	5040
gctttgggga	caccatattc	tccccattcc	tactttcctc	aatggccctc	ctcagtctcc	5100
tttggaagaa	ggaaaaagaa	acttcattat	ctcctggatg	tagtacaac	aactcaagct	5160
caacatgtgc	atactgaact	ccatttcctt	ttcccaaact	tcgacattta	cagccatccc	5220
ctttcagctg	atagcaagtt	tatccttcca	gctactcaaa	ccagaatctt	tagagccatc	5280
cttgaccctt	tctctcctct	cacactcaac	atctatccat	cagaaaattt	tgttggttct	5340
actttcaaaa	tgcatacaga	gtcagagcat	gtctcattac	ctccaatagc	taccatacta	5400
gtctgaacaa	acatcatttc	tcacctgggt	tattgaacaa	acatcatttc	tcacctgggt	5460
tattgatagc	atcctaacgg	gtcttcctgt	ttcttggttc	ccctatatta	gcaacacagc	5520
agtcagagga	gtccttttag	aactcaatca	gatcatgtca	cgctactcct	ctacttaaaa	5580
tccttcaatg	ggtcccatta	cacaaagagt	acaaaccaga	gcccttacac	tggctctacaa	5640

5/11

gttccaacat	ttgactcctg	ttatctctct	gacatcatat	tctaataatta	ctgctgttgt	5700
cctttttgctc	cagtcacact	gtttgattag	taaatattta	ttaaacaaag	caatcctagt	5760
ctccaaagag	atcatagttt	attggaggaa	acaagagcct	ataaatgggt	acacacagaa	5820
ggtagtgatt	atggttctcc	ctcacctccc	atcctaaact	ttgacagggt	aaactccct	5880
ggatgttgaa	ggttgaggaa	tttgccaggg	ttcaggggtg	tggtggagga	ggcagggagg	5940
aagcaaggac	atttcaggca	ggaagaacat	tacatgcaaa	gatctaaaga	tatgaatcag	6000
caacatattt	atggaattac	aagtaaagta	gaaagttc			6038

<210> 3  
 <211> 542  
 <212> DNA  
 <213> Homo sapiens

<400> 3						
tcactgttag	caagtaactg	actttataga	ccaatatgcc	tctcttctga	aatgggtctta	60
ttttaaacia	atgtgagcaa	aagaaaatat	ttatgagatt	ctaaaaatga	agacataatt	120
ttgtagtata	gaattttctt	ggccaggaat	ggtggctcat	gcttgtaatc	ccagcacttt	180
gggaggccaa	ggtcagagga	ttgcttgagc	ctggaagggt	gaagatgcag	tgattcatga	240
ttataccact	gcactccagc	ctgggcaaca	gagcaagacc	ctgtctcaag	aaaagaaaag	300
aatttttattt	ttcttttctc	acaaaaatag	actttaaaat	aataatggaa	gaacaaatat	360
gatgatcaca	attatcagag	taattacttt	atgacagtca	gcaataagat	tctaactctt	420
aaatattcct	ctgcttaaat	catttatattg	gagttttgat	ctataatata	ttcccaccct	480
gacccaaaaa	ttgaagaagg	acaaggaaaa	atgttggttc	aagaaacaaa	gatgtaagta	540
aa						542

<210> 4  
 <211> 3213  
 <212> DNA  
 <213> Homo sapiens

<400> 4						
actaacataa	agctgaaggt	gaataaaaaa	atcaggggtta	gccaaacaaa	ttttcatggt	60
caaataccac	ataaaaagta	aatatactta	agttcccagc	aaaatctgaa	ttgaacgtag	120
acaaaatgct	cattttctcag	tgtttgacag	acttaacagt	ttgagccaat	aaaaatgtac	180
tgactagata	aactactaaa	agttgttaat	ttttgcaatg	tatatctctg	aaaagaaagt	240
ttatctatta	tagaaattcc	tgtgcccatt	taagaacttt	gagcatttta	attgtttaat	300
aatatagttt	aattgcatca	tgaaaataat	caataataca	atttatttgg	tttattttaa	360
aaaactgatt	ctttctgctc	tctctatata	tagactgatt	ttataactaa	gttgccctaa	420
gatcaccaaa	ttgtttgaag	cctagggttc	tgagggatgg	aaaatgatgt	cacaactatt	480
tacagttcac	acacacattc	tggggattta	atacatcctt	tacaagtgca	ggaaagggtg	540
aagattgatg	atgtggggga	attagagcta	ccacacccca	gaggggtgga	tggtatgttg	600
tctgttggtg	gctgtgtgaa	tcagagaggt	tgatttagac	atatatttag	aaagaggaaa	660
gatgaaccaa	tcaaaaataa	taactataat	gacttttcaa	gatatagaca	atacagttaa	720
gatataaatg	gaaacaaaaa	aagttaaaag	tggggagatg	aagtctgatt	ttttgggttt	780
tttttttttt	tgcttttttg	tttggttatg	taatcagtg	taccagttta	aaataatggg	840
ttataagaca	ctatatgcaa	gcctcatggt	aacctccaat	ctaaaacata	caacaaatac	900
acacaaaata	aaaaggagaa	attaaaacac	accaccagag	aaaatcacct	acattaaaag	960
aaagacaaat	aggaagaaaa	taagaaagag	aaggccatca	aataatcaga	aatgaataa	1020
caaatgaca	ggaataagtc	ctcataaata	ataacattga	atgtaaattg	actaagctct	1080
ccaatgaaag	acagggagtg	gctgaatgta	ttttaaaaaa	aatattacac	cgagctgtgc	1140
gtggtgtctc	acacctataa	tcccagcatt	ttgggagact	gagccgggtg	gatcacttga	1200
gccagaggat	tcgagaccag	cctggccaac	atggcaaaaac	cctgtctcta	ctaaaaatac	1260
aaaaaattag	ctgaacatgg	tggcacatgc	ctgtgggttc	agctactaga	gaggctgagg	1320
cagaagaatt	gcttgaactt	gggaggtgga	ggttgcagtg	agctaagatt	gatggagcca	1380
ctgcacccca	gcctagggtga	cagaataaga	ctctgcctca	aaaaaaaaaa	gcaaaacaaa	1440
acaaaacaaa	aaacccttag	acccaatgat	tcattgccta	caagaagtat	gcttcacctt	1500
taaagacaca	tatagactga	aggtaaaggg	atggaaaaat	attctatgcc	tatggaaaca	1560
aacaaaaaga	agcagaagct	acattttatat	cagacaaaat	agactgcaag	acaaaaacta	1620
tgaaaagaga	gaaagaaggt	cattatatag	tgataaaggg	gtccatttag	caagagcatt	1680
taacaattct	aaatatatat	tcacccaata	ctggagtact	caggtatata	aagcaaatat	1740
tattagagcc	aaagagagag	atagacagac	ccccatacaa	taataactgg	agacttcaac	1800
accccacttt	cagcattgga	cagatcatcc	agacagaaaa	ttaacaaaca	tcaaatttca	1860



6/11

tctgcacccat	aggtcaaatg	gacctagtag	atattttacag	aacattttgat	ccaacagctg	1920
tagaatacac	atttcttctcc	tcagcacatg	gataattctc	aaggatatac	caaagtctag	1980
gtcacaaaaac	aaatctttaa	atttagaaaa	aaagtgaat	aatatcaaac	gttttctctc	2040
accacagact	aagaaaaaaa	gaagtcccaa	ataaatacaa	tctgagataa	aaaaggagac	2100
gagacaacca	ataccacaaa	aaattaaagg	atcattagaa	gatactatga	aactatatgc	2160
taataaattg	gaaaacctga	acaaaataga	taattcctag	aaacatacaa	catactgggc	2220
tgttcaggtt	ttgtattttt	tcatagtacc	atgaagaaat	acaagaattg	tttctagaac	2280
cattcttgta	tttcttcatg	gtttttgtat	ttcttcatgg	aaccatgaag	aaatacaaaa	2340
tgtgaacagg	ccaataacaa	gtaatgagac	agaagccata	ctaaaaagta	tcccagaaaa	2400
gaactcagga	tctgatggct	tcactgatga	attttgccaa	atatttaaaa	aactaatacc	2460
aatccaactc	aaattattaa	aaaaatagag	gtggacagaa	tctttccaaa	tgtattctat	2520
gaggccagtg	ttttttctga	ttgaatctcc	cattatatatt	taatcacata	taaaaccaga	2580
gaaagacaca	ttaaaaagaa	agaaaactgt	aggccaatat	ctctgatgaa	cattgatgca	2640
gaaatcctca	acaacaaatt	agcaaaactga	attcaagaac	acattaaaac	aatcattcat	2700
catgaccaag	tgggaattgt	cctagagatt	caagtgtggt	taggtatgtg	cagatcaatg	2760
ggtttaatgt	tgtccaatga	acataatgtc	ctccagctcc	atccatgttc	ttgcaaataga	2820
caggatctca	ttctttttta	tggctaagta	gtactccatt	gtgtataagt	gccatatttt	2880
ctttatccat	tcattctgtta	gacacctaag	ttgcttccaa	atcttagcta	ttgtgaatag	2940
tgctgcaata	aacatgggag	tgtaaataatt	ttgttgacat	actgatttca	tttcctttgg	3000
ataaataccc	agtagtgga	ttgctggatc	atatggggga	aaatggagat	ggctaacggg	3060
ctcaaaaata	tagttagaaa	aatgaatat	gatttagtat	tcgatagcac	aataggatga	3120
ctactgttaa	tgataattta	ttatatatta	taaaataact	aaaatagtat	aatgggatg	3180
tatgtagcag	agagaaatga	taaatgtttg	aag			3213

&lt;210&gt; 5

&lt;211&gt; 6679

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 5

gtcgacctgc	aggtcaacgg	atcacttgag	gacagtagtt	caagaccagc	ctgggcagca	60
tagggagact	gtctctacga	aaaatcaaaa	aattatggcc	gggcatggtg	gctcacgtct	120
gtaatccctg	aactttggga	catcaaggca	agtggatcac	ttgagggtcag	gagttcgaga	180
ctagcctggc	caacatgggtg	aaaccctatc	tccactaaaa	aatacaaaaa	ttagccaggc	240
atgggtggcag	gcacctgtaa	tcccgggtac	tcaggaggct	gaggcaggag	aatcacttga	300
accagggagg	cggagggttg	agtgagctga	gatcacacca	ctgcactcca	gcctgggtga	360
cagagcaaga	ctctatctca	aaaaaaataa	aaaaataaaa	aaattagcca	ggcatggtag	420
tgcacacctc	tagtctcagc	tactcaggag	gctgagggtg	gaggatcact	tgaacctggg	480
gcagtcaagg	ctacagtga	ccaagatcat	gccactacac	tccagcctgg	gcaacagaga	540
gagaccctgt	ctctaaaaaa	ataataataa	taaagaaaaa	aacagctctg	tttatgtctc	600
ctgggtccata	catactacta	tgtatatagt	ttgcaaactc	aaagatccag	atagtcaatt	660
tttttaggctt	gtgggcccga	tgggtctctgt	cacaatcact	ctgccctgtc	tttctagcac	720
aaaagcagct	ataaacaata	catacatgaa	tttttttatag	acatcgagat	ttgaatttca	780
tatgattttt	acattttata	aaataatctt	tttaaaaaatt	ttcccctaac	catttataaag	840
tgtaaaagcc	ggccagcgcg	ccatcgtcac	gcctgttaatt	ccagcacttt	gggaggctga	900
ggtgggcaga	tcacttgaga	tcaacagttc	gagaccagcc	tggccaacat	agcaaaaccc	960
catttctact	aaaaataaaa	aaattagctg	ggcatagtgg	tgcacacctg	tgatcccagc	1020
tacttgggag	gctgaggcag	gagaatcgct	tgaacctggg	aagcggagggt	tgcagtgagc	1080
caacatcatg	ccactgcact	ccagcctggg	tgacagagtg	agacttcgtc	tcaacgaaaa	1140
aaaaaagtgt	aaaagccatt	cctaattcag	tgtacatcag	tgtacatact	caggtctgcg	1200
tactcctgct	ctgaggcata	cctgagaagt	agagttgctt	ggtcacagga	catacacatt	1260
tccacattaa	ctagacacta	ccaagttgcc	atccaaggag	gttttttttt	tacaatctac	1320
actcccccca	gcaacaaatg	agagttactc	cagatccctt	acaaagatgc	tctaagccca	1380
gtaccagatg	aaaacaggaa	gtgggagggg	aagctgccag	ccccttctaa	ccatgaagaa	1440
atacctggta	gagccttctg	gatgctggaa	ggatgaataa	cgggggtctc	tggagcctgc	1500
cccctgtcag	atcactgtga	cttctgagcc	tccagtcag	tctcagcccc	atgtgtcatg	1560
gccagtgata	atgagccctc	actctctgtt	tgggtctttat	tctccccatg	tggggctgaa	1620
gtctggattg	agccgttatt	caagatgtac	agctttcttg	acaggaaagt	agtgtcacag	1680
aaacagcagg	ggcttggcaa	gatgatctaa	ctgcaaatcc	tacctggctc	agccaccagc	1740
tagttctgtg	atcttgaaca	agttttttca	cttctctgag	gccatccctt	ggctacaaca	1800
caccagttgg	ttgacaggat	gaaatgacga	agtcccttac	acctgtaatc	ccagcacttt	1860
gggaggccaa	ggcgggtgga	tggcttgagc	ctgagagggtg	acagcatgcc	ggcagtcctc	1920

7/11

acagccctcg	ttcgcctctcg	gcgcctcctc	tgcctgggct	cccacttcgg	tggcacttga	1980
ggagcccttc	agcccaccgc	tgcactgtgg	gagccctttt	ctgggctggc	caaggccaga	2040
gccggtccc	tcagcttgca	gggaggtgtg	gagggagagg	ctcaagcagg	aaccggggct	2100
gcgcacggcg	cttgccgggc	agctggagtt	ccgggtgggc	gtgggcttgg	cgggccccgc	2160
actcggagca	gcggggccagc	cctgccaggc	cccgggcaat	gagaggctta	gcacccgggc	2220
cagcggctgc	ggaggggtgt	ctgggtgccc	cagcagtgcc	agcccggcgg	cgctgtgctc	2280
gctcgatttc	tcactggggc	ttagcagcct	tcccgcgggg	cagggtcggg	gacctgcagc	2340
ccgccatgcc	tgagcctccc	ctccatgggc	tcctgtgcgg	cccagacctc	cccagcagac	2400
accacccctt	gctccacagc	gccagtcctc	atcgaccacg	caagggtctga	gaagtgcggg	2460
cgcacggcac	cgggactggc	aggcagctac	ccctgcagcc	ctgggtgcga	atccactggg	2520
tgaagccagc	tgggctcctg	agtctgggtg	agacttggag	aacctttatg	tctagctcag	2580
ggatcgtaaa	tacaccaatc	agcaccctgt	gtctagctca	gggtctgtga	atgcaccaat	2640
ccacactctg	tatctagcta	ctctgatggg	gccttggaga	acctttatgt	ctagctcagg	2700
gattgtaaat	acaccaatcg	gcactctgta	tctagctcaa	ggtttgtaaa	cacaccaatc	2760
agcaccctgt	gtctagctca	gggtatgtga	atgcaccaat	cgacagtctg	tatctggcta	2820
ctttcatggg	catccgtgtg	aagagaccac	caaacaggct	ttgtgtgagc	aataaagctt	2880
ctatcacctg	ggtgcagggtg	ggctgagtc	gaaaagagag	tcagcgaagg	gagataaggg	2940
tggggccggt	ttataggatt	tgggtaggt	aaggaaaatt	acagtcaaag	ggggtttgtt	3000
ctctggcggg	caggagtggg	gggtcgcaag	gtgctcagtg	ggggtgcttt	ttgagccagg	3060
atgagccagg	aaaaggactt	tcacaaggta	atgtcatcaa	ttaaggcaag	gacccgccat	3120
ttacacctct	tttgtgggtg	aatgtcatca	gttaagtgtg	ggcagggcac	attcacttct	3180
tttgtgattc	ttcagttact	tcaggccatc	tgggcgtata	tgtgcaagtt	acaggggatg	3240
cgatggcttg	gcttgggctc	agaggcttga	cagctactct	ggtggggcct	tggagaatgt	3300
ttgtgtcgac	actctgtatc	tagttaatct	agtggggacg	tggagaacct	ttgtgtctag	3360
ctcagggatt	gtaaacgcac	caatcagcgc	cctgtcaaaa	cagaccactc	ggctctacca	3420
atcagcagga	tgtgggtggg	gccagataag	agaataaaaag	caggctgccc	gagccagcag	3480
tggcaacgcg	cacagggtccc	tatccacaat	atggcagctt	tgttcttttg	ctgtttgcga	3540
taaatcttgc	tactgctcgc	tttttgggtc	cacactgctt	ttatgagctg	taacactcac	3600
cacgaagggtc	tgcagcttca	ctcctgaagc	cactaagacc	acgagcccac	cgggaggaat	3660
gaacaactcc	ggccgcgctg	ccttaagagc	tataacactc	accgcgaagg	tctgcagctt	3720
cactcctcag	ccagcgagac	cacgaaccca	ccagaaggaa	gaaactgcga	acacatctga	3780
acatcagaag	gaacaaactc	cagatgcacc	accttaagag	ctgtaacact	cactgcgagg	3840
gtccgcggct	tccttcttga	agtcagttag	accaagcact	caccagtttc	ggacacaagc	3900
ccaggagtth	gagatcagcc	tgggcaacat	gatgaaatgc	cctctctgca	aaaaaaaaaa	3960
aaattacaaa	aattggcggg	gcattggtgt	ccgtgcctgt	ggtcccagct	acgcgggagg	4020
ctaaagtggg	aggatcgctt	gagcctggga	ggtgaagact	gcagttagct	gtgattgtac	4080
cacagccctc	taggctgggg	gacagactga	gacctgtttt	cccctccgca	aaaaaattga	4140
caaaagtgt	ataagagggtg	cctgatatgg	ctaggcgcag	tggctcatgc	ctgtaatccc	4200
agcacttttg	gaagccgagg	cgggcgggtc	acctaaagtc	aggagtgtga	gaccagcctg	4260
gccaacatgg	agaaagccca	tctcttctaa	aaatacaaaa	ttagccggct	gtgggggcag	4320
tgggtggagca	tgcctgtaat	cccagctact	caggaggctg	aggcaggaga	atcacttgaa	4380
cccaggaggc	ggcgggttga	gtgagccgag	atcgtgccat	tgcactccac	ccactccagc	4440
ctgggcaaca	agagccaaac	tctgtcttaa	aaaaaaaaaa	aaaaagtgcc	tgacatataa	4500
gaggtgtgca	atgcaatagt	tgccaggcaa	catgtttaag	aatgtggagc	tcctgccttc	4560
catggctcctg	ttaaaaaccc	acctcaagg	ccagggtgcag	tggctcatgc	ctataatccc	4620
agcacttttg	gaggccgagg	cgggtggatc	acctgaggtc	aggagtctga	gaccagcctg	4680
accaccaaca	tgggtgaaatc	ccacctctac	taaaaatata	aaattagatg	agcatgggtg	4740
tgcattgcctg	taatcccacc	tacttgggag	gctgaggcag	gaaaatcact	agaaccaggg	4800
aggcggagggt	tgtagtgagc	cgagatcgtg	ccattgcact	ccagcctgag	caatgagcga	4860
aactccatct	caaaaaaaca	acaacaaaaa	cccactctct	actcccaggg	agctgggtac	4920
agagctgggc	cacatcagtg	caagggtgctg	agccacagag	ctaaggcgga	gctgcaggac	4980
cgcggaccag	ataacagtgt	gtgagatcag	tgtgtgagat	cagacgtccc	tgccattgggt	5040
gaccaccagg	gggcccccaa	gcaccagaga	tggccccatc	cagtcaccac	atccacttct	5100
catccagaga	tgtctgtttc	ttggcacgct	ggggtaaatt	aggacagaag	gtgacagtct	5160
tgggtgtggt	cagtccagact	gccccaggca	ggccttgtgg	cctgtagaaa	acgttcaggc	5220
ctaggccggg	cacgggtggct	cacgcctgta	atcccagcac	tttgggaggc	cgaggcgggt	5280
ggatcacgag	gtcaggagat	cgtgaccatc	ctgggtataa	cgggtgaaacc	ccgtctctac	5340
taaaaatata	aaaaattggc	cgggcattggt	ggcgggcacc	tgtagtcca	gctactcggg	5400
aggctgaggc	aggagaatgg	cgtgaacccg	agaggcagag	tttgcagtga	gccgagatcg	5460
cgcactgca	ctccagcctg	ggcgacagag	caagactcca	tctggaaaag	aaaaagaaaa	5520
cgttcaggtc	tgagccagag	gcccaggctg	taattctgtc	acttaccatg	accttgggca	5580
aggcacttcc	ttccctggcc	cagttcacgg	ggttggaatc	gactccaagg	tccttccag	5640



8/11

cattaacgct	gcatgggttct	aagatgagaa	gatggggcag	tttccccctct	ctcacccccag	5700
cccgtgtcca	cttcaaggtg	aatgaccagg	gaagtcacgt	gtcccaatcc	cgcagttcca	5760
aagcccttgg	ggaccctact	gtcagggctcg	tgcacgagga	ggtgaaggtc	aggtgagcca	5820
atcgctcga	agggtcttgc	ctcattcggg	acagacatcc	ggtttcctct	ggctctaccg	5880
ggattctagg	ggcttttagcc	gaatgagtca	tggggggcgg	gggggtttct	gggggagttc	5940
ccagctaata	aacttgggac	aggacagcct	ggaactttcg	atgggtgccta	tccaagtgtg	6000
gggtggggcac	agcagccaag	acccaatgtc	cttatctcag	gtaggggctc	aggaggtctc	6060
ccagacaggc	agcctccgga	gagtttgggg	gtaggaatgg	gagcaaccag	gcttcttttt	6120
ttctctctta	gaatttgggg	gcttggggga	caggcttgag	aatcccaaag	gagaggggca	6180
aaggacactc	ccccacaagt	ctgccagagc	gagagaggga	gaccccgact	cagctgccac	6240
ttccccacag	gcctctgccc	cttccaggcg	tctatcagcg	gctcagcctt	tgttcagctg	6300
ttctgttcaa	acactctggg	gccattcagg	cctgggtggg	gcagcgggag	gaagggagtt	6360
tgaggggggc	aaggcgacgt	caaaggagga	tcagagattc	cacaatttca	caaaactttc	6420
gcaaacagct	ttttgttcca	acccccctgc	attgtcttgg	acaccaaatt	tgcataaatc	6480
ctgggaagtt	attactaagc	cttagtcgtg	gccccaggta	atttcctccc	aggcctccat	6540
ggggttatgt	ataaaggggc	ccctagagct	ggggcccaaa	acagcccgga	gcctgcagcc	6600
cagccccacc	cagacccatg	gctggacctg	ccaccagag	cccatgaag	ctgatgggtg	6660
agtgtcttgg	cccaggatg					6679

&lt;210&gt; 6

&lt;211&gt; 6235

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 6

gatcacttga	ggacagtagt	tcaagaccag	cctggggcagc	atagggagac	tgtctctacg	60
aaaaatcaaa	aaattatggc	cgggcatggg	ggctcacgtc	tgtaatccct	gaactttggg	120
acatcaaggc	aagtggatca	cttgagggtca	ggagttcgag	actagcctgg	ccaacatggg	180
gaaaccctat	ctccactaaa	aaatacaaaa	attagccagg	catggtggca	ggcacctgta	240
atccccgcta	ctcaggaggc	tgaggcagga	gaatcacttg	aacccaggag	gcggagggtg	300
cagtgaagctg	agatcacacc	actgcactcc	agcctgggtg	acagagcaag	actctatctc	360
aaaaaaaaata	aaaaaataaa	aaaatttagcc	aggcatggta	gtgcacacct	ctagtctcag	420
ctactcagga	ggctgagggtg	ggaggatcac	ttgaacctgg	ggcagtcacg	gctacagtga	480
gccaaagatca	tgccactaca	ctccagcctg	ggcaacagag	agagaccctg	tctctaaaaa	540
aataataata	ataaagaaaa	aaacagctct	gtttatgtct	cctgggtccat	acatactact	600
atgtatatag	tttgcaaaact	caaagatcca	gatagtcaat	tttttaggct	tgtggggccgt	660
atgggtctctg	tcacaatcac	tctgccctgt	ctttctagca	caaaagcagc	tataaacaat	720
acatacatga	atTTTTTTata	gacatcgaga	tttgaatttc	atatgatttt	tacatttttat	780
aaaataatct	tttttaaaaat	tttccccctaa	ccatttaaaa	gtgtaaaagc	cggccagcgc	840
gccatcgta	cgctgtaat	tccagcactt	tgggaggctg	aggtggggcag	atcacttgag	900
atcaacagtt	cgagaccagc	ctggccaaca	tagcaaaacc	ccatttctac	taaaaataaa	960
aaaattagct	gggcatagtg	gtgcacacct	gtgatcccag	ctacttggga	ggctgaggca	1020
ggagaatcgc	ttgaacctgg	gaagcggagg	ttgcagttag	ccaacatcat	gccactgcac	1080
tccagcctgg	gtgacagagt	gagacttcgt	ctcaacgaaa	aaaaaaagtg	taaaagccat	1140
tcctaattca	gtgtacatca	gtgtacatac	tcagggtctgc	gtactcctgc	tctgaggcat	1200
acctgagaag	tagagttgct	tggtcacagg	acatacacat	ttccacatta	actagacact	1260
accaagttgc	catccaagga	ggtttttttt	ttacaatcta	cactcccccc	agcaacaaat	1320
gagagttact	ccagatcctt	tacaaagatg	ctctaagccc	agtaccagat	gaaaacagga	1380
agtgggaggg	gaagctgcc	gccccttcta	accatgaaga	aatacctggg	agagccttct	1440
ggatgctgga	aggatgaata	acgggggtct	ctggagcctg	ccccctgtca	gatcactgtg	1500
acttctgagc	ctccagtcga	gtctcagccc	catgtgtcat	ggccagtgat	aatgagccct	1560
cactctctgt	ttgggtcttta	ttctccccat	gtggggctga	agtctggatt	gagccgttat	1620
tcaagatgta	cagcttttctt	gacaggaaag	tagtgtcaca	gaaacagcag	gggcttggca	1680
agatgatcta	actgcaaata	ctacctggct	cagccaccag	ctagttctgt	gatcttgaac	1740
aagtttttttc	acttctctga	ggccatccct	tggctacaac	acaccagttg	gttgacagga	1800
tgaatatgacg	aagtccctta	cacctgtaat	cccagcactt	tgggaggcca	aggcgggtgg	1860
atggcttgag	cctgagaggt	gacagcatgc	cggcagtcct	cacagccctc	gttcgctctc	1920
ggcgcctcct	ctgcctgggc	tcccacttcg	gtggcacttg	aggagccctt	cagcccaccg	1980
ctgcactgtg	ggagcccctt	tctgggctgg	ccaaggccag	agccggctcc	ctcagcttgc	2040
agggaggtgt	ggagggagag	gctcaagcag	gaaccggggc	tgcgcacggc	gcttgcgggc	2100
cagctggagt	tccgggtggg	cgtgggcttg	gcggggcccc	cactcggagc	agcggggccag	2160
ccctgccagg	ccccgggcaa	tgagaggctt	agcaccgggg	ccagcggctg	cggagggtgt	2220

10/11

caggacagcc	tggaactttc	gatggtgcct	atccaagtgt	ggggtgggca	cagcagccaa	6000
gacccaatgt	ccttatctca	ggtaggggct	caggaggtct	cccagacagg	cagcctccgg	6060
agagtttggg	ggtaggaatg	ggagcaacca	ggcttctttt	tttctctctt	agaattttggg	6120
ggcttggggg	acaggcttga	gaatcccaaa	ggagaggggc	aaaggacact	ccccacaag	6180
tctgccagag	cgagagaggg	agaccccgac	tcagctgcca	cttccccaca	ggcct	6235

<210> 7  
 <211> 278  
 <212> DNA  
 <213> Homo sapiens

<400> 7						
aagcttttat	aggtgtaa	tttccactta	gtactgcttt	tgtaatggtg	tcttttttatt	60
ttcattttat	tcaagatggt	ttctaatttc	tcttgacttc	cttcttaaat	tcttacctca	120
tgtagacata	cattttttggc	cctatgcatt	gggatgcaaa	accagactaa	tttactttgt	180
acaaaaagaa	aatgagaaa	gaaatatatt	tggtcttggtg	agcactatat	ggaaataactt	240
tatattccat	ttgtttcatc	atattcatat	atcccttt			278

<210> 8  
 <211> 73  
 <212> DNA  
 <213> Homo sapiens

<400> 8						
cattggatac	tccatcacct	gctgtgatat	tatgaatgtc	tgccatatata	aatattcact	60
attccataac	aca					73

<210> 9  
 <211> 3033  
 <212> DNA  
 <213> Homo sapiens

<400> 9						
actaacataa	agctgaaggt	gaataaaaaa	atcagggtta	gccaaacaaa	ttttcatggt	60
caaataccac	ataaaaagta	aatatactta	agttcccagc	aaaatctgaa	ttgaacgtag	120
acaaaatgct	cattttctcag	tgtttgacag	acttaacagt	ttgagccaat	aaaaatgtac	180
tgactagata	aactactaaa	agttgtta	ttttgcaatg	tatatattctg	aaaagaaagt	240
ttatctatta	tagaaattcc	tgtgcccatt	taagaacttt	gagcatttta	attgtttaat	300
aatatagttt	aattgcatca	tgaaaataat	caataatata	atttatattgg	tttatattaaa	360
aaaactgatt	ctttctgctc	tctctatata	tagactgatt	ttataactaat	gttgcctaaa	420
gatcaccaaa	ttgtttgaag	cctaggtttc	tgagggatgg	aaaatgatgt	cacaactatt	480
tacagttcac	acacacattc	tggggattta	atacatcctt	tacaagtgca	ggaaagggtg	540
aagattgatg	atttggggga	attagagcta	ccacacccca	gagggtggta	tggtatgttg	600
tctgttggtg	gctgtgtgaa	tcagagagtt	tgatttagac	atatatttag	aaagaggaaa	660
gatgaaccaa	tcaaaaataa	taactataat	gacttttcaa	gatatagaca	atacagttaa	720
gatataaatg	gaaacaaaaa	aagttaaaaa	tggggagatg	aagtctgatt	ttttgggtttt	780
tttttttttt	tgcttttttg	tttgtttatg	taatcagtgt	taccagttta	aaataatggg	840
ttataagaca	ctatatgcaa	gcctcatggt	aacctccaat	ctaaaacata	caacaaatac	900
acacaaaata	aaaaggagaa	attaaaacac	accaccagag	aaaatcacct	acattaaaag	960
aaagacaaat	aggaagaaaa	taagaaagag	aaggccatca	aataatcaga	aaatgaataa	1020
caaatgaca	ggaataagtc	ctcataaata	ataacattga	atgtaaatgg	actaagctct	1080
ccaatgaaag	acagggagtg	gctgaatgta	ttttaaaaaa	aatattacac	cgagctgtgc	1140
gtggtgtctc	acacctataa	tcccagcatt	ttgggagact	gagccgggtg	gatcacttga	1200
gcccaggagt	tcgagaccag	cctggccaac	atggcaaaac	cctgtctcta	ctaaaaatac	1260
aaaaaattag	ctgaacatgg	tggcacatgc	ctgtggttcc	agctactaga	gaggctgagg	1320
cagaagaatt	gcttgaactt	gggaggtgga	ggttgcatgt	agctaagatt	gatggagcca	1380
ctgcacccca	gcctaggtga	cagaataaga	ctctgcctca	aaaaaaaaaa	gcaaaaacaaa	1440
acaaaacaaa	aaacccttag	acccaatgat	tcattgccta	caagaagtat	gcttcacctt	1500
taaagacaca	tatagactga	aggtaaaggg	atggaaaaat	attctatgcc	tatggaaaca	1560
aacaaaaaga	agcagaagct	acatttatat	cagacaaaat	agactgcaag	acaaaaacta	1620
tgaaaagaga	gaaagaaggt	cattatatag	tgataaaggg	gtccatttag	caagagcatt	1680
taacaattct	aaatatatat	tcacccaata	ctggagtact	caggtatata	aagcaaatat	1740



9/11

actgggtgcc	ccagcagtgc	cagccccgccg	gcgctgtgct	cgctcgattt	ctcactgggc	2280
cttagcagcc	ttcccgcggg	gcagggctcg	ggacctgcag	cccgccatgc	ctgagcctcc	2340
cctccatggg	ctcctgtgcg	gcccagagcct	ccccgacgag	caccaccccc	tgctccacag	2400
cgcccagtcc	catcgaccac	gcaagggctg	agaagtgcgg	gcgcacggca	ccgggactgg	2460
caggcagcta	cccctgcagc	cctgggtgcgg	aatccactgg	gtgaagccag	ctgggctcct	2520
gagtctgggtg	gagacttgga	gaacctttat	gtctagctca	gggatcgtaa	atacaccaat	2580
cagcaccctg	tgtctagctc	agggctctgtg	aatgcaccaa	tccacactct	gtatctagct	2640
actctgatgg	ggccttgagg	aacctttatg	tctagctcag	ggattgtaaa	tacaccaatc	2700
ggcactctgt	atctagctca	aggtttgtaa	acacaccaat	cagcaccctg	tgtctagctc	2760
agggatatgtg	aatgcaccaa	tgcacagtct	gtatctggct	actttcatgg	gcatccgtgt	2820
gaagagacca	ccaaacaggc	tttgtgtgag	caataaagct	tctatcacct	gggtgcaggt	2880
gggctgagtc	cgaaaagaga	gtcagcgaag	ggagataagg	gtggggccgt	tttataggat	2940
ttgggtaggt	aaaggaaaat	tacagtcaaa	gggggtttgt	tctctggcgg	gcaggagtgg	3000
ggggctcgcaa	ggtgctcagt	gggggtgctt	tttgagccag	gatgagccag	gaaaaggact	3060
ttcacaaggt	aatgtcatca	attaaggcaa	ggacccgcc	tttacacctc	ttttgtgggtg	3120
gaatgtcatc	agttaagttg	gggcagggca	tattcacttc	ttttgtgatt	cttcagttac	3180
ttcaggccat	ctgggcgtat	atgtgcaagt	tacaggggat	gcgatggctt	ggcttgggct	3240
cagaggcttg	acagctactc	tgggtggggcc	ttggagaatg	tttgtgtcga	cactctgtat	3300
ctagttaatc	tagtggggac	gtggagaacc	tttgtgtcta	gctcagggat	tgtaaacgca	3360
ccaatcagcg	ccctgtcaaa	acagaccact	cggtcttacc	aatcagcagg	atgtgggtgg	3420
ggccagataa	gagaataaaa	gcaggctgcc	cgagccagca	gtggcaacgc	gcacaggtec	3480
ctatccacaa	tatggcagct	ttgttctttt	gctgtttgcg	ataaatcttg	ctactgctcg	3540
ctttttgggt	ccacactgct	tttatgagct	gtaacactca	ccacgaaggt	ctgcagcttc	3600
actcctgaag	ccactaagac	cacgagccca	ccgggaggaa	tgaacaactc	cggccgcgct	3660
gccttaagag	ctataaacact	caccgcgaag	gtctgcagct	tcactcctca	gccagcgaga	3720
ccacgaaccc	accagaagga	agaaaactgcg	aacacatctg	aacatcagaa	ggaacaaact	3780
ccagatgcac	caccttaaga	gctgtaaacac	tcactgcgag	ggtccgcggc	ttccttcttg	3840
aagtcagtga	gaccaagcac	tcaccagttt	cggacacaag	cccaggagtt	tgagatcagc	3900
ctgggcaaca	tgatgaaatg	ccctctctgc	aaaaaaaaaa	aaaattacaa	aaattggcgg	3960
agcatggtgg	tccgtgcctg	tgggtcccagc	tacgcgggag	gctaaagtgg	gaggatcgct	4020
tgagcctggg	aggtgaagac	tgcagtgagc	tgtgattgta	ccacagccct	ctaggctggg	4080
ggacagactg	agaccctgtt	tccctctcgc	aaaaaaattg	acaaaagtgt	aataagaggt	4140
gcctgatatg	gctaggcgca	gtggctcatg	cctgtaatcc	cagcactttg	ggaagccgag	4200
gcgggcgggt	cacctaaaggt	caggagtgtg	agaccagcct	ggccaacatg	gagaaagccc	4260
atctcttcta	aaaatacaaa	attagccggc	tgtgggggca	gtggtggagc	atgcctgtaa	4320
tcccagctac	tcaggaggct	gaggcaggag	aatcacttga	accagggagg	cggcggttgc	4380
agtgagccga	gatcgtgcc	ttgcactcca	cccactccag	cctgggcaac	aagagccaaa	4440
ctctgtctta	aaaaaaaaaa	aaaaaagtgc	ctgacatata	agaggtgtgc	aatgcaatag	4500
ttgccaggca	acatgtttaa	gaatgtggag	ctcctgcctt	ccatggtcct	gttaaaaacc	4560
cacctcaag	gccaggtgca	gtggctcatg	cctataatcc	cagcactttg	ggaggccgag	4620
gcgggtggat	cacctgaggt	caggagtctg	agaccagcct	gaccaccaac	atggtgaaat	4680
cccacctcta	ctaaaaatac	aaaattagat	gagcatggtg	gtgcatgcct	gtaatcccac	4740
ctacttgga	ggctgaggca	ggaaaatcac	tagaaccagg	gaggcggagg	ttgtagttag	4800
ccgagatcgt	gccattgcac	tccagcctga	gcaatgagcg	aaactccatc	tcaaaaaaac	4860
aacaacaaaa	acccactctc	tactcccagg	gagctgggta	cagagctggg	ccacatcagt	4920
gcaaggtgct	gagccacaga	gctaaggcgg	agctgcagga	ccgcggacca	gataacagtg	4980
tgtgagatca	gtgtgtgaga	tcagacgtcc	ctgccattgg	tgaccaccag	ggggccccc	5040
agcaccagag	atggccccat	ccagtcacca	catccacttc	tcacccagag	atgtctgttt	5100
cttggcacgc	tggggtaaat	taggacagaa	ggtgacagtc	ttgggtgtgg	tcagtcagac	5160
tgccccaggc	aggccttgtg	gcctgtagaa	aacgttcagg	cctaggccgg	gcacggtggc	5220
tcacgcctgt	aatcccagca	ctttgggagg	ccgaggcggg	tggatcacga	ggtcaggaga	5280
tcgtgaccat	cctggctaac	acggtgaaac	ccgtctcta	ctaaaaatac	aaaaaattgg	5340
ccgggcatgg	tggcgggcac	ctgtagtctc	agctactcgg	gaggctgagg	caggagaatg	5400
gcgtgaaccc	gagaggcaga	gtttgcagtg	agccgagatc	gcgccactgc	actccagcct	5460
gggcgacaga	gcaagactcc	atctggaaaa	gaaaaagaaa	acgttcagggt	ctgagccaga	5520
ggcccaggct	gtaattctgt	cacttaccat	gaccttgggc	aaggcacttc	cttccttggc	5580
ccagttcacg	gggttggaat	cgactccaag	gtcccttcca	gcattaacgc	tgcattggtc	5640
taagatgaga	agatggggca	gtttcccttc	tctcaccoca	gcccggtgtc	acttcaaggt	5700
gaatgaccag	ggaagtacag	tgtcccaatc	ccgcagttcc	aaagcccttg	gggaccctac	5760
tgtcagggtc	gtgcacgagg	aggtgaaggt	caggtgagcc	aatcgctcgc	aagggtcttg	5820
cctcattcgg	gacagacatc	cggtttcttc	tggctctacc	gggattctag	gggcttttag	5880
cgaatgagtc	atggggggcg	gggggggtttc	tgggggagtt	cccagcta	caacttggga	5940

11/11

tattagagcc	aaagagagag	atagacagac	ccccatacaa	taataactgg	agacttcaac	1800
accccacttt	cagcattgga	cagatcatcc	agacagaaaa	ttaacaaaca	tcaaatttca	1860
tctgcacccat	aggtcaaagt	gacctagtag	atattttacag	aacatttgat	ccaacagctg	1920
tagaatacac	attctttctcc	tcagcacatg	gataattctc	aaggatatac	caaagtctag	1980
gtcacaaaaac	aatctttaa	atttagaaaa	aaagtgaat	aatatcaaac	gttttctctc	2040
accacagact	aagaaaaaaa	gaagtcccaa	ataaatacaa	tctgagataa	aaaaggagac	2100
gagacaacca	ataccacaaa	aaattaaagg	atcattagaa	gatactatga	aactatatgc	2160
taataaattg	gaaaacctga	acaaaataga	taattcctag	aaacatacaa	catactggtc	2220
tgttcaggtt	ttgtattttt	tcatagtacc	atgaagaaat	acaagaattg	tttctagaac	2280
cattcttgta	tttcttcatg	gtttttgtat	ttcttcatgg	aaccatgaag	aaatacaaaa	2340
tgtgaacagg	ccaataacaa	gtaatgagac	agaagccata	ctaaaaagta	tcccagaaaa	2400
gaactcagga	tctgatggct	tcactgatga	atthttgccaa	atattttaaaa	aactaatacc	2460
aatccaactc	aaattattaa	aaaaatagag	gtggacagaa	tctttccaaa	tgtattctat	2520
gaggccagtg	ttttttctga	ttgaatctcc	cattatattt	taatcacata	taaaaccaga	2580
gaaagacaca	ttaaaaagaa	agaaaactgt	aggccaatat	ctctgatgaa	cattgatgca	2640
gaaatcctca	acaacaaatt	agcaaactga	attcaagaac	acattaaaac	aatcattcat	2700
catgaccaag	tggaatttgt	cctagagatt	caagtgtggt	taggtatgtg	cagatcaatg	2760
ggtttaaatgt	tgtccaatga	acataatgtc	ctccagctcc	atccatgttc	ttgcaaatga	2820
caggatctca	ttctttttta	tggctaagta	gtactccatt	gtgtataagt	gccatatttt	2880
ctttatccat	tcatctgtta	gacacctaag	ttgcttccaa	atcttagcta	ttgtgaatag	2940
tgctgcaata	aacatgggag	tgtaaatatt	ttgttgacat	actgatttca	tttcctttgg	3000
ataaataccc	agtagtgagg	ttgctggatc	ata			3033